



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 164274

TO: Richard Schnizer
Location: rem/2D30/2C18
Art Unit: 1635
Thursday, September 08, 2005
Case Serial Number: 09/910432

From: Barb O'Bryen
Location: Biotech-Chem Library
Remsen 1a69
Phone: 571-272-2518

BOB
barbara.obryen@uspto.gov

Search Notes

This Page Blank (uspto)

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

^o
This Page Blank (uspto)

RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 [3]

RP SEQUENCE FROM N.A.
 RA Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
 RA Maves H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [4]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL022141; CAA18129.1; -

DR EMBL; AL161589; CAB80294.1; -

DR PIR; T04592; T04592.

DR InterPro; IPR002952; Eggshell.

DR PRINTS; PR01228; EGGSHLL.

SQ SEQUENCE 221 AA; 20344 MW; DE8DAE2E2C57A9F8 CRC64;

Query Match 74.8%; Score 223; DB 2; Length 221;

Best Local Similarity 55.8%; Pred. No. 6.1e-11;

Matches 43; Conservative 1; Mismatches 7; Indels 26; Gaps 1;

Qy 1 GGGGGGGG-----GGGGGGGGGGRDRRRRRRRGGG 34

Db 96 GGGGGGGGGGSSGGGGGGGNGKDNKSHRNKSSGGGGGGGGGGNGSGRGGG 155

Qy 35 GGGGGGGGGGGGGGGGGG 51

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

Qy 156 GGGGGGGGGGGGGGGGGG 172

Db 156 GGGGGGGGGGGGGGGGGG 172

CAZ_DROME

ID -CAZ_DROME STANDARD; PRT; 399 AA.

AC Q27294; Q24445; Q8T3M2; Q9VXI4;

DT 16-OCT-2001 (Rel. 40, Created)

DT 05-JUL-2004 (Rel. 44, Last sequence update)

DT 25-JAN-2005 (Rel. 46, Last annotation update)

DE RNA-binding protein cabeza (Sarcoma-associated RNA-binding fly

DE homolog) (P19).

GN Name=caz; Synonym=SAARFH; ORFNames=CG3606;

OS Drosophila melanogaster (Fruit fly);

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS A AND B), FUNCTION, SUBCELLULAR LOCATION,

RP DEVELOPMENTAL STAGE, AND TISSUE SPECIFICITY.

RC STRAIN=Canton-S;

RX MEDLINE=95349623; PubMed=7623847;

RA Immanuel D., Zinsner H., Ron D.;

RT "Association of SARFH (sarcoma-associated RNA-binding fly homolog)

RT with regions of chromatin transcribed by RNA polymerase II.";

RL Mol. Cell. Biol. 15:4562-4571(1995).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM A), AND TISSUE SPECIFICITY.

RC STRAIN=Canton-S;

RX MEDLINE=95223793; PubMed=7708500;

RA Stelow D.T., Haynes S.R.;

RT "Cabeza, a Drosophila gene encoding a novel RNA binding protein,

RT shares homology with EWS and TUS, two genes involved in human sarcoma

RT formation.";

RL Nucleic Acids Res. 23:835-843(1995).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;

RA Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos J.D.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Fesler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou S., Zhu X., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

Science 287:2185-2195(2000).
 [4]
 GENOME REANNOTATION, AND ALTERNATIVE SPLICING.
 MEDLINE=22426069; PubMed=12537572;
 Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochownik S.E.,
 Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 Betencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
 Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 Stapleton M., Yamada C., Ashburner M., Gelbart W.W., Rubin G.M.,
 Lewis S.E.;
 "Annotation of the Drosophila melanogaster euchromatic genome: a
 systematic review";
 Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 [5]
 SEQUENCE FROM N.A. (ISOFORM B).
 STRAIN=Berkeley; TISSUE=Embryo;
 Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.W.,
 Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 George R.A., Gonzalez M., Guarin H., Kronmiller B., Li P.W., Liao G.,
 Miranda A., Mungall C.J., Nunoo J., Pacleb J.M., Faragas V., Park S.,
 Patel S., Phouanavong S., Wan K.H., Yu C., Lewis S.E., Rubin G.M.,
 Celniker S.E.;
 Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
 [6]
 SEQUENCE OF 39-404 FROM N.A.
 STRAIN=Oregon-R;
 Haynes S.R.;
 Submitted (APR-1988) to the EMBL/GenBank/DBJ databases.
 [7]
 SEQUENCE OF 212-261 FROM N.A.
 STRAIN=Oregon-R;
 MEDLINE=8717568; PubMed=3031652;
 Haynes S.R., Rebert M.L., Moxer B.A., Forquignon F., Dawid I.B.;
 "Pen repeat sequences are GGN clusters and encode a glycine-rich
 domain in a Drosophila cDNA homologous to the rat helix destabilizing
 protein";
 Proc. Natl. Acad. Sci. U.S.A. 84:1819-1823(1987).
 CC -!- FUNCTION: May participate in a function common to the expression
 of most genes transcribed by RNA polymerase II.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=B;
 CC IsoId=Q27294-1; Sequence=Displayed;
 CC Name=A;
 CC IsoId=Q27294-2; Sequence=VSP_005778;
 CC Note=No experimental confirmation available;
 CC -!- TISSUE SPECIFICITY: Ubiquitous. Enriched in the brain and central
 nervous system during embryogenesis. Enriched in the adult head.
 CC Embryos contain both isoforms A and B, whereas later in
 development (heads and torsos) only isoform B is detected.
 CC -!- DEVELOPMENTAL STAGE: Expressed in the developing embryo from the
 earliest stages of cellularization and is subsequently found in
 many cell types.
 CC -!- MISCELLANEOUS: 'Cabeza' means 'head' in Spanish.
 CC -!- SIMILARITY: Belongs to the RNP TET family.
 CC -!- SIMILARITY: Contains 1 RanBP2-type zinc finger.
 CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; U13178; AAA86955.1; -;
 EMBL; L37083; AAC41563.1; -;
 EMBL; AB003501; AAF48578.2; -;
 EMBL; AB003501; AAN09389.1; -;

DR EMBL; AY094763; AAM11116.1; ALT_SEQ.
 DR EMBL; BT004875; AAC45231.1; -;
 DR EMBL; M15765; AAA70425.1; -;
 DR PIR; S54729; S54729.
 DR HSP; O95218; INOZ.
 DR FlyBase; FBgn0011571; caz.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR InterPro; IPR001876; Znf_RanGDP.
 DR Pfam; PF00076; RRM_1; 1.
 DR Pfam; PF00641; zf-RanBP; 1.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00547; Znf RBZ; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS01358; ZF_RANBP2_1; 1.
 DR PROSITE; PS50199; ZF_RANBP2_2; 1.
 KW Alternative splicing; Metal-binding; Nuclear protein; RNA-binding;
 KW Zinc-finger.
 FT DOMAIN 4 393 Gly-rich.
 FT DOMAIN 120 206 RNA-binding (RRM).
 FT ZN_FING 275 304 RanBP2-type.
 FT VARSPDIC 4 47 Missing (in isoform A).
 FT CONFLICT 39 41 /FTId=VSP_005778.
 FT CONFLICT 92 92 PNY -> LFI (in Ref. 4).
 FT CONFLICT 109 109 H -> P (in Ref. 1, 2 and 6).
 FT CONFLICT 254 254 Missing (in Ref. 1, 2 and 6).
 FT CONFLICT 278 278 G -> GNGGGGG (in Ref. 1, 2 and 6).
 FT CONFLICT 384 393 D -> E (in Ref. 6 and 7).
 FT CONFLICT 384 393 DGGPMRDGG -> MVDQKRWS (in Ref. 6).
 SQ SEQUENCE 399 AA; 38839 MW; 3001FC96BDD6FDCA CRC64;
 Query Match 69.3%; Score 206.5; DB 1; Length 399;
 Best Local Similarity 68.9%; Pred. No. 2.1e-09;
 Matches 42; Conservative 0; Mismatches 2; Indels 17; Gaps 3;
 Qy 1 GGGGGGGG-----GGGGGGGGGGGGRDRRRRRRRGGGGGGG---GGGGGGGGGG 48
 Db 215 GGGGGGGGGGFGGRRGGGGGGGGGGGGGGRFD-----RGGGGGGRYDRGGGGGGGGG 269
 Qy 49 G 49
 Db 270 G 270
 RESULT 5
 Q18444
 ID Q18444 PRELIMINARY; PRT; 136 AA.
 AC Q18444;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein C34D4.11.
 GN Name=C34D4.11; ORFNames=C34D4.11;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=90069613; PubMed=9851916;
 RG WormBase Consortium;
 RT "Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z., Le T.T.;
 RT "The sequence of C. elegans cosmid C34D4.";
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;

RA Zhou L., Fraser C.M.;
 RT "Structural flexibility in the Burkholderia mallei genome.";
 RL Proc. Natl. Acad. Sci. U.S.A. 101:14247-14251(2004).
 DR EMBL; CP000010; AAU48078.1; -.
 SQ SEQUENCE 199 AA; 19605 MW; 390222C03692C8D89 CRC64;

Query Match 68.1%; Score 203; DB 2; Length 199;
 Best Local Similarity 68.9%; Pred. No. 2.2e-09;
 Matches 42; Conservative 0; Mismatches 9; Indels 10; Gaps 3;

Qy 1 GGGGGGGGGGGG--GGGGGGGRRDRRRRRRRGGGG--GGGG--GGGGGGGGGG 50
 Db 116 GGGGGGGGGGGGDDGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 175

Qy 51 G 51
 Db 176 G 176

RESULT 9
 Q9U2U0 PRELIMINARY; PRT; 285 AA.

AC Q9U2U0;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein Y116A8C.35.
 GN Name=uaf-2; ORFNames=Y116A8C.35;
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=95069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA McMurray A.A.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL117204; CAB55137.1; -.
 DR PIR; T31503; T31503.
 DR HSP; Q01081; 1UMT.
 DR InAct; Q9U2U0; -.
 DR WormBase; WBGene0000698; uaf-2.
 DR WormPep; Y116A8C.35; C233341.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR InterPro; IPR009106; CART_C.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR InterPro; IPR009145; U2_small.
 DR InterPro; IPR000571; Znf_CCCH.
 DR Pfam; PF00076; RRM 1; 1.
 DR Pfam; PF00642; Zf-CCCH; 2.
 DR PRINTS; PR01848; UZAUFACTOR.
 DR SMART; SM00360; RRM; 1.
 DR SMART; SM00356; Znf_C3H1; 2.
 DR PROSITE; PS0102; RRM; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 285 AA; 30856 MW; E2B932629653C040 CRC64;

Query Match 68.1%; Score 203; DB 2; Length 285;
 Best Local Similarity 76.5%; Pred. No. 3e-09;
 Matches 39; Conservative 0; Mismatches 8; Indels 4; Gaps 1;

Qy 1 GGGGGGGGGGGG--GGGGGGGRRDRRRRRRRGGGGGGGGGGGGGGGGGGGG 51
 Db 218 GGG 264

RESULT 10

Q8QKX8 PRELIMINARY; PRT; 698 AA.

AC Q8QKX8;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE EsV-1-144.
 GN Name=ORF 144;
 OS Ectocarpus siliculosus virus.
 OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phaeovirus.
 OX NCBI_TaxID=37665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EsV-1;
 RX MEDLINE=20192171; PubMed=10725207; DOI=10.1006/viro.2000.0225;
 RA Delaroque N., Wolf S., Muller D.G., Knippers R.;
 RT "Characterization and immunolocalization of major structural proteins
 RT in the brown algal virus EsV-1.";
 RL Virology 269:148-155(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EsV-1;
 RA Delaroque N., Bothe G., Pohl T., Knippers R., Mueller D.G., Boland W.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF204951; AAP28325.1; -.
 SQ SEQUENCE 698 AA; 74035 MW; C7638DDB922E0D50 CRC64;

Query Match 67.8%; Score 201.5; DB 2; Length 698;
 Best Local Similarity 68.4%; Pred. No. 8.2e-09;
 Matches 39; Conservative 0; Mismatches 3; Indels 15; Gaps 2;

Qy 1 GGGGGGGGGG-----GGGGGGGGGRRDRRRRRRRGGGGGGGGGGGGGGGG 51
 Db 454 GGGGGGGGTGGAGSGAGSGGGGGGTGGAG-----GGGGGGGGGGGGGGGG 501

RESULT 11

Q95UW6 PRELIMINARY; PRT; 161 AA.

AC Q95UW6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE No on or off transient A (Fragment).
 GN Name=nonA;
 OS Drosophila ezoana (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=47313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=E20;
 RA Huttunen S., Vieira J., Hoikkala A.;
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY012600; AAG48870.1; -.
 DR FlyBase; FBgn0046740; Dezo\nonA.
 FT NON TER 1 1
 FT NON TER 161 161
 SQ SEQUENCE 161 AA; 15708 MW; 28A68346D17AB988 CRC64;

Query Match 67.4%; Score 201; DB 2; Length 161;
 Best Local Similarity 77.1%; Pred. No. 2.7e-09;
 Matches 37; Conservative 0; Mismatches 7; Indels 4; Gaps 1;

Qy 1 GGGGGGGGGGGG--GGGGGGGRRDRRRRRRRGGGGGGGGGGGGGGGGGGGG 48
 Db 71 GGG 114

Search completed: September 7, 2005, 12:59:57
Job time : 113.5 secs

This Page Blank (uspto)

(I) is useful for identifying a compound that modulates taste signaling in taste cells which involves contacting the compound with (I) and determining the functional effect of the compound (II). The functional effect is determined by measuring changes in intracellular cAMP, cGMP, IP₃, or Ca²⁺. The functional effect is a chemical or physical effect. The functional effect is determined by measuring binding of the compound to an extracellular domain of (I). The functional effect is determined by measuring binding of radiolabeled GTP to (I). (I) is recombinant and is from rat, mouse, or human. (I) is expressed in a cell or cell membrane. The functional effect is measured by determining changes in the electrical activity of cells (e.g., eukaryotic cell) expressing (I). (I) is useful for identifying compound that modulates taste signaling in taste cells which involves contacting a compound with (I). The extracellular domain (II) of (I) is useful for identifying a compound that modulates taste signaling in taste cells which involves contacting the compound with the polynucleotide encoding (I) and determining the functional effect of the compound upon the extracellular domain. This is the amino acid sequence of a T2R G-protein coupled receptor related linker.

```

SQ Sequence 200 AA;
Query Match      76.7%; Score 228.5; DB 8; Length 200;
Best Local Similarity 80.4%; Pred. No. 1.5e-14;
Matches 41; Conservative 0; Mismatches 1; Indels 9; Gaps 1;

Qy 1 GGGGGGGGGGGGGGGGGGGGGGGGGRKKRQRRRGGGGGGGGGGGGGGGGGGGG 51
Db 1 GGGGGGGGGGGGGGGGGGGGGGGG-----GGGGGGGGGGGGGGGGGGGG 42

```

RESULT 6	
ADM96215	
ID	ADM96215 standard; protein; 200 AA.
XX	
XX	ADM96215;
XX	
XX	
DT	01-JUL-2004 (first entry)
XX	
XX	Poly Gly flexible linker.
XX	
KW	G-protein coupled receptor 39; GPR 39; cancer; breast; ovary; prostate;
KW	brain; lung; ligand binding; aberrant receptor protein analysis;
KW	cytostatic; receptor; poly Gly flexible linker.
XX	
OS	Synthetic.

Key	Location/Qualifiers
Key	6. .200
FT	Misc-difference
FT	5. .200
FT	/note= "Residues are independently optionally absent"
XX	
XX	
PN	US2004071708-A1.
XX	
XX	
PD	15-APR-2004.
XX	
XX	
PF	26-SEP-2002; 2002US-00255775.
XX	
XX	
PR	26-SEP-2002; 2002US-00255775.
XX	
XX	
PA	(IMMU-) IMMUSOL INC.

PI Claassen G, Li H, Barber J;
XX
XX WPI: 2004-373880/35.
XX
XX
XX Identifying anticancer agents modulating G-protein coupled receptor
PT proteins, comprises contacting GPR 39-specific binding agents to cancer
PT cells and detecting anticancer activity to identify anticancer agent.
XX
XX
PS Disclosure; SEQ ID NO 3: 33pp; English.
XX
XX The invention relates to a method of identifying anticancer agents that
CC modulate G-protein coupled receptor (GPR) proteins, comprising contacting

GPR 39-specific binding agents to cancer cells and detecting anticancer activity. The invention also relates to a method of inhibiting cancer characteristics in cancer cells by down modulating GPR 39 protein activity to a level sufficient to inhibit the cancer characteristics of the cancer cells and an antibody that specifically recognises a GPR 39 protein. The method further involves binding a population of different compositions to the GPR 39 protein to select GPR 39-specific binding agents. The cancer cells are chosen from breast, ovarian, prostate, brain and lung cancer cells. The method is useful for identifying anticancer agents that modulate GPR 39 protein and for inhibiting cancer characteristics in cancer cells. The antibody is useful for inhibiting receptor function e.g. by blocking ligand binding, which is used in therapeutic function, in various in vitro histochemical techniques for detecting GPR proteins and as a diagnostic tool for aberrant receptor protein analysis. This sequence represents a poly Gly flexible linker used in the method of the invention.

Sequence 200 AA;

Query Match	76.7%	Score 228.5;	DB 8;	Length 200;
Best Local Similarity	80.4%	Pred. No. 1.5e-14;		
Matches 41;	Conservative	0;	Mismatches 1;	Indels 9;
Gaps	17;			

[illegible]

RESULT 7
ABG71303
ID ABG71303 standard; protein; 201 AA.

AC	ABG71303;
XX	
DT	17-DEC-2002 (first entry)
XX	
DE	PINPOINT poly-Gly linker protein.
XX	
KW	Protein position identification with nuclease tail; linker; PINPOINT; molecular beacon.
KW	

OS	Synthetic.		
XX			
FH	Key	Location/Qualifiers	
FT	Misc-difference	1. .97	/note= "Optionally absent"
FT			105. .201
FT	Misc-difference		/note= "Optionally absent"

PN	US6444421-B1.
XX	
XX	
PD	03-SEP-2002.
XX	
XX	
PF	02-APR-1998; 98US-00054281.
XX	
XX	
PR	03-APR-1997; 97US-0080234P.
PR	03-APR-1997; 97US-00825664.
XX	
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Chung JH;
XX WPI; 2002-749290/81.
DR

PT Detecting whether a first molecule is in close proximity to a second
PT molecule, or detecting target molecule, by using a molecular beacon
PT comprising an oligonucleotide to which a fluorophore and a quencher are
PT attached.

PS Disclosure; Col 59-60; 41pp; English.

The invention relates to detecting (M) whether a first molecule is in close proximity to a second molecule, or detecting a target molecule, CC
CC
CC

CC involves using a molecular beacon that comprises an oligonucleotide (ON)
CC to which a fluorophore and a quencher (FQ) are attached. The method is
CC named Protein Position Identification with Nuclease Tail (PINPOINT). The
CC method further comprises, whether a first molecule is in close proximity
CC to a second molecule involves attaching a molecular beacon to the first
CC molecule, where the molecular beacon comprises an (ON) to which is
CC attached an (FQ), attaching an endonuclease (EN) moiety to the second
CC molecule, and determining whether the first molecule is in close
CC proximity to the second molecule by detecting whether fluorescence is
CC emitted by the fluorophore, where fluorescence emission is indicative of
CC cleavage of the (ON) by the (EN) moiety, to cause separation of the (FQ).
CC The method may also comprise, detecting a target molecule, which involves
CC contacting the target molecule with a chimaeric fusion molecule that
CC comprises an (EN) molecule and a guide molecule that binds to the target
CC molecule, thus guiding the chimaeric fusion molecule to the target
CC molecule, contacting the chimaeric endonuclease with a molecular beacon
CC comprising an (ON) to which is attached a fluorophore and a quencher, and
CC detecting the presence of a fluorescent signal which results from
CC cleavage of the (ON) by the (EN), to allow separation of the quencher
CC from the fluorophore. The method is useful for detecting whether a first
CC molecule is in close proximity to a second molecule, or for detecting a
CC target molecule and is useful for detecting intermolecular interactions
CC in vivo, in situ and in vitro, such as enzymatic reactions, hormone-
CC ligand interactions, and drug or toxin interactions. The present sequence
CC is a proteinlinker used in the PINPOINT constructs of the invention
XX
XX
XX Sequence 201 AA:
SQ

```
Query Match      76.7%   Score 228.5; DB 5; Length 201;
Best Local Similarity 80.4%; Pred. No. 1.5e-14;
Matches 41; Conservative 0; Mismatches 1; Indels 9;
```

RESULT 8	
ABU22369	
ID	ABU22369 standard; protein; 188 AA.
XX	
XX	ABU22369;
XX	
XX	
DT	19-JUN-2003 (first entry)
XX	
DE	Protein encoded by Prokaryotic essential gene #7896.
XX	
KW	Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX	
OS	Burkholderia mallei.
XX	
PN	WO200277183-A2.
PN	
PD	03-OCT-2002.
XX	
XX	
PF	21-MAR-2002; 2002WO-US009107.
XX	
PR	21-MAR-2001; 2001US-00815242.
PR	06-SEP-2001; 2001US-00948993.
PR	25-OCT-2001; 2001US-0342923P.
PR	08-FEB-2002; 2002US-00072851.
PR	06-MAR-2002; 2002US-0362699P.
XX	
PA	(ELIT-) ELITRA PHARM INC.
XX	
PI	Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI	Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX	
DR	WPI; 2003-025926/02.
DR	N-PSDB; ACA26239.
XX	
PT	New antisense nucleic acids, useful for identifying proteins or screening
PT	for homologous nucleic acids required for cellular proliferation to

isolate candidate molecules for rational drug discovery programs.

Claim 25; SEQ ID NO 50293; 1766pp; English.

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 188 AA:

```

Query Match          72.0%; Score 214.5; DB 6; Length 188;
Best Local Similarity 67.2%; Pred. No. 3.le-13;
Matches 43; Conservative 2; Mismatches 6; Indels 13; Gaps 3
Qy    1  GGGGGGGGGGGGGG-----GGGGGGYGRKRRRGGG---GGGG-----GGGGGGGGGG 47
      |||||
Db    115 GGGGGGGGGGGGGDGDGYGGGGGCGGGRDWRGGGGGRASGGGGAGARSGGGGGGGG 174
      |||||
Qy   48 GGGG 51
      ||||
Db   175 GGGG 178
      ||||

```

RESULT 9	
ABP29010	
ID	ABP29010 standard; protein; 56 AA.
XX	
XX	
AC	ABP29010;
XX	
XX	
02-JUL-2002	(first entry)
DT	
XX	
DE	Streptococcus polypeptide SEQ ID NO 7196.
XX	
XX	
KW	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW	group A streptococcus; Streptococcus pyogenes; antibacterial;
KW	antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX	
OS	Streptococcus agalactiae.
XX	
PN	WO200234771-A2.
XX	
XX	
PD	02-MAY-2002.
XX	
PF	29-OCT-2001; 2001WO-GB004789.
XX	
XX	


```

PN WO200171042-A2.
XX 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US009231.
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL08832.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 20979; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 344 AA;
XX
XX Query Match 66.8%; Score 199; DB 4; Length 344;
XX Best Local Similarity 65.1%; Pred. No. 1.5e-11;
XX Matches 41; Conservative 1; Mismatches 5; Indels 16; Gaps 3;
XX
Qy 1 GGGGGGGGGG-----GGGGGGGGGGYGRKRRRQRRGGG-----GGGGGGGGGGGG 48
Db 12 GGGGGGGGGGPRGGGGGGGGGGFG-----GGRGGGGGGGDRGGGGGGGGGG 67
XX
Qy 49 GGG 51
Db 68 GGG 70
XX
RESULT 12
ABB60010
ID ABB60010 standard; protein; 399 AA.
XX
AC ABB60010;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 6822.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
FD 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX 11-JUL-2000; 2000US-00614150.
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL08832.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 6822; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 399 AA;
XX
XX Query Match 66.6%; Score 198.5; DB 4; Length 399;
XX Best Local Similarity 67.2%; Pred. No. 1.9e-11;
XX Matches 41; Conservative 0; Mismatches 3; Indels 17; Gaps 3;
XX
Qy 1 GGGGGGGGGG-----GGGGGGGGGGYGRKRRRQRRGGGGGGG-----GGGGGGGGGGGG 48
Db 215 GGGGGGGGGGFRGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 269
XX
Qy 49 G 49
Db 270 G 270
XX
RESULT 13
ABO59421
ID ABO59421 standard; protein; 208 AA.
XX
XX ABO59421;
AC ABO59421;
XX
DT 29-JUL-2004 (first entry)
XX
DE Human genome derived single exon protein #5655.
XX
KW Human; gene expression; single exon probe; microarray;
KW alternative splicing event; genomic alteration.
XX
OS Homo sapiens.
XX
PN US2003194704-A1.
XX
PD 16-OCT-2003.
XX
PF 03-APR-2002; 2002US-00029386.
XX
XX 03-APR-2002; 2002US-00029386.
XX
PA (PENN/) PENN S G.
PA (RANK/) RANK D R.
PA (HANZ/) HANZEL D K.
XX
PI Penn SG, Rank DR, Hanzel DK;
XX WPI; 2004-119264/12.
XX
XX New human genome-derived single exon nucleic acid probes useful for human
XX gene expression analysis, for identifying or characterizing alternative
XX splicing events, for assessing genomic alterations or as tools for
XX surveying tissues.

```


PI Zoghbi HY, Van Den Veyver IB, Amir R, Francke U;

WPI: 2004-256068/24.

PT Screening human for Rett syndrome comprises detecting mutation in nucleic acid sequence encoding methyl-CpG-binding protein 2 (MECP2)

acid sequence encoding methyl-CpG-binding protein 2 (MECP2).

PS Disclosure; SEQ ID NO 112; 125pp; English.

The invention relates to a method of screening a human for Rett syndrome comprising detecting a mutation in a nucleic acid sequence encoding methyl-CpG-binding protein 2 (MECP2). The method is useful for screening a human for Rett syndrome. The method is useful for screening neurodevelopmental diseases such as Rett syndrome, autism, non-syndromic mental retardation, idiopathic neonatal encephalopathy, idiopathic infantile spasms, idiopathic cerebral palsy, Angelman syndrome and schizophrenia. The present sequence represents the amino acid sequence of a methyl-CpG-binding protein.

The invention relates to a method of screening a human for Rett syndrome comprising detecting a mutation in a nucleic acid sequence encoding

comprising detecting a mutation in a nucleic acid sequence encoding methyl-CpG-binding protein 2 (MECP2). The method is useful for screening

methy1-CpG-binding protein 2 (MECP2): the method is useful for screening a human for Rett syndrome. The method is useful for screening

neurodevelopmental diseases such as Rett syndrome, autism, non-syndromic a human for Rett syndrome. The method is useful for screening

neurodevelopmental diseases such as Rett syndrome, autism, non-syndromic mental retardation, idiopathic neonatal encephalopathy, idiopathic

infantile spasms, idiopathic cerebral palsy, Angelman syndrome and

intensive spasms, paraparesis, paraplegia, angelman syndrome and schizophrenia. The present sequence represents the amino acid sequence of

a methyl-CpG-binding protein.

Sequence 345 AA;

Query Match 65.4%; Score 195; DB 8; Length 345;

Best Local Similarity 34.6%; Pred. No. 3.6e-11;

Matches	44;	Conservative	2;	Mismatches	5;	Indels	76;	Gaps	2;
---------	-----	--------------	----	------------	----	--------	-----	------	----

Qy

1 GGGGGGGGGGGG--GGGYGR----- 23

D_b 197 GGGGGGGGGGGGGRVCGGGGGRVRAAAERGGGRLLVKMPFAGGAPASPPPTLP 256

0v 24 -----KRRRORRRGGGGGGGGGG 44

Db 257 PSAHPPPTAPPATTHGOGI GGGVKT PCRKRKA EANDSPVPKRKRKP GCGCGCGCGGGG 316

0v 45 CCCCCC 51

317 323

Search completed: September 7, 2005, 12:56:08
Job time : 66.5 secs

Job time : 66.5 secs

This Page Blank (uspto)

Result No.	Query		DB	ID	Description
	Score	Match Length			
1	228.5	76.7	200	4	US-09-589-981A-13
2	228.5	76.7	201	3	US-09-052-995-1
3	228.5	76.7	201	3	US-09-053-003-40
4	228.5	76.7	201	3	US-09-054-281-22
5	228.5	76.7	201	4	US-09-478-948-6
6	228.5	76.7	201	4	US-09-618-094-40
7	228.5	76.7	201	4	US-09-754-947-5
8	195	65.4	345	4	US-09-657-013-112
9	194	65.1	162	3	US-09-575-574-4
10	192.5	64.6	462	4	US-09-319-039-324
11	189.5	63.6	526	4	US-09-538-092-1080
12	183.5	61.6	112	4	US-08-643-681A-8319
13	183	61.4	268	2	US-08-635-099A-9
14	183	61.4	268	3	US-09-157-349-9
15	181.5	60.9	60	4	US-09-832-297A-12
16	181.5	60.9	100	4	US-08-411-067C-4
17	181.5	60.9	738	3	US-08-864-038A-3
18	178.5	59.9	68	4	US-09-107-433-4787
19	178.5	59.9	111	4	US-09-543-681A-8322
20	178	59.7	266	3	US-09-032-523-7
21	178	59.7	266	4	US-09-802-633-7
22	177.5	59.6	179	4	US-09-270-767-57837
23	176.5	59.2	40	3	US-08-815-190A-17
24	176.5	59.2	40	4	US-09-333-213-1
25	173	58.1	1958	1	US-07-945-283-2
26	171.5	57.6	334	4	US-09-949-016-11006
27	171	57.4	204	4	US-09-107-433-4573

RESULT 3
US-09-053-003-40
; Sequence 40, Application US/09053003
; Patent No. 6207391
; GENERAL INFORMATION:
; APPLICANT: Wu, Pengguang
; APPLICANT: McKinney, Judi
; TITLE OF INVENTION: High-Throughput Screening Assays for
; TITLE OF INVENTION: Modulators of STAT4 and STAT6 Activity
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/053_003
; FILING DATE: 31-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367

RESULT 4
US-09-054-281-22
; Sequence 22, Application US/09054281
; Patent No. 644421
; GENERAL INFORMATION:
; APPLICANT: Chung, Jay H.
; TITLE OF INVENTION: Methods for Detecting Intermolecular
; TITLE OF INVENTION: Interactions In Vivo and In Vitro
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/054,281
; FILING DATE: 02-APR-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/080,234
; FILING DATE: 03-APR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/826,622
; FILING DATE: 03-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Timothy L.
; REGISTRATION NUMBER: 35,367
; REFERENCE/DOCKET NUMBER: 015280-2952000US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO. 22:

[illegible]

```
; FILE REFERENCE: 014907-003310US
; CURRENT APPLICATION NUMBER: US/09/754,947
; CURRENT FILING DATE: 2001-01-04
; PRIOR APPLICATION NUMBER: US 60/174,901
; PRIOR FILING DATE: 2000-01-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: flexible linker
; NAME/KEY: MOD RES
; LOCATION: (1)..(97)
; OTHER INFORMATION: Gly at positions 1-97 may be present or absent
; NAME/KEY: MOD RES
; LOCATION: (105)..(201)
; OTHER INFORMATION: Gly at positions 105-201 may be present or absent
US-09-754-947-5

Query Match          76.7%; Score 228.5; DB 4; Length 201;
Best Local Similarity 80.4%; Pred. No. 2.3e-15;
Matches 41; Conservative 0; Mismatches 1; Indels 9; Gaps 1;

QY 1 GGGGGGGGGGGGGGGGGGGGGGVRKKRRRRRRGGGGGGGGGGGGGGGGGG 51
Db 1 GGGGGGGGGGGGGGGGGGGGGG-----GGGGGGGGGGGGGGGGGGGG 42

RESULT 8
US-09-657-013-112
; Sequence 112, Application US/09657013
; Patent No. 6709817
; GENERAL INFORMATION:
; APPLICANT: Zoghbi, Huda Y.
; APPLICANT: Van den Veyver, Ignatia B
; APPLICANT: Amir, Ruthie
; APPLICANT: Francke, Uta
; TITLE OF INVENTION: Methods of Identifying Mutations in a Methyl-CPG-Binding Domain
; FILE REFERENCE: HO-P01893US1/0905371
; CURRENT APPLICATION NUMBER: US/09/657,013
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: US 60/152,778
; PRIOR FILING DATE: 1999-09-07
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 112
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Chicken
US-09-657-013-112

Query Match          65.4%; Score 195; DB 4; Length 345;
Best Local Similarity 34.6%; Pred. No. 6e-12;
Matches 44; Conservative 2; Mismatches 5; Indels 76; Gaps 2;

QY 1 GGGGGGGGGGGGGGGGGGGGGGGGGYGR-----GGGGYGR----- 23
Db 197 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 256
QY 24 -----KKRRRRRRGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 44
Db 257 PSAAHPPTAPPATHGGGLGGGVKRPGRKKAEDSRVPKRRKPGGGGGGGGGGG 316
QY 45 GGGGGGGG 51
Db 317 GGGGGVGG 323

RESULT 9
US-09-575-574-4
```

```
; Sequence 4, Application US/09575574
; Patent No. 6376750
; GENERAL INFORMATION:
; APPLICANT: Yu, Su-May
; APPLICANT: Chao, Yu-Chan
; TITLE OF INVENTION: PLANT SEEDLING AND EMBRYO PROMOTER
; FILE REFERENCE: 08919-047001
; CURRENT APPLICATION NUMBER: US/09/575,574
; CURRENT FILING DATE: 2000-05-22
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 162
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-575-574-4

Query Match          65.1%; Score 194; DB 3; Length 162;
Best Local Similarity 65.0%; Pred. No. 4e-12;
Matches 39; Conservative 0; Mismatches 5; Indels 16; Gaps 2;

QY 1 GGGGGGGGGGGGG-----GGGGGGYGRKKRRRRRRGGGGGGGGGGGGGG 50
Db 90 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 143

RESULT 10
US-09-919-039-324
; Sequence 324, Application US/09919039
; Patent No. 6727066
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 324
; LENGTH: 462
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6727066 181344CD1
US-09-919-039-324

Query Match          64.6%; Score 192.5; DB 4; Length 462;
Best Local Similarity 48.2%; Pred. No. 1.3e-11;
Matches 41; Conservative 3; Mismatches 6; Indels 35; Gaps 3;

QY 1 GGGGGGGGGGG-----GGG-----GGGGGGYGRKKRRRRRRGGGGGG 38
Db 166 GGGGGGGGGGGYQDQSSMSGGGGGGGGYQDQSGGGGGGGYQDQDRGRRRGGGGGG 225
QY 39 GGGGGG-----GGGGGG 50
Db 226 GGGGGGYNRSSGGYEPGRGGGRGG 250

RESULT 11
US-09-538-092-1080
; Sequence 1080, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
```

```
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 1080
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P35637
; US-09-538-092-1080

Query Match      63.6%; Score 189.5; DB 4; Length 526;
Best Local Similarity 48.2%; Pred. No. 2.9e-11;
Matches 41; Conservative 2; Mismatches 7; Indels 35; Gaps 3;

Qy 1 GGGGGGGGGG-----GGG-----GGGGGGYGRKKRRQRRRGGGGGG 38
Db 166 GGGGGGGGGYQDQSSMSGGGGGYGNQDQSGGGGGYGGQDRGRGRGGGGGG 225

Qy 39 GGGGG-----GGGGG 50
Db 226 GGGGGGYNRRSGGYEPRGGRGG 250

RESULT 12
US-09-543-681A-8319
; Sequence 8319, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8319
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-8319

Query Match      61.6%; Score 183.5; DB 4; Length 112;
Best Local Similarity 69.2%; Pred. No. 3e-11;
Matches 36; Conservative 1; Mismatches 4; Indels 11; Gaps 2;

Qy 1 GGGGGGGGGGGGGG--GGGGYGRKKRRQRRRGGGGGGGGGGGGGGGGGGGG 50
Db 5 GGGGGGGGGGGGGGGGGGGGKKXGGG--RGRGRGGGGGGGGGGGGGG 47

RESULT 13
US-08-835-099A-9
; Sequence 9, Application US/08835099A
; Patent No. 5874277
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasushi
; APPLICANT: NISHI, Kazuhiro
; APPLICANT: KAWAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; TITLE OF INVENTION: AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston

; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 1080
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P35637
; US-09-538-092-1080

Query Match      61.4%; Score 183; DB 2; Length 268;
Best Local Similarity 73.9%; Pred. No. 6.9e-11;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 6 GGGGGGGGGGGGGGGGGGGYGRKKRRQRRRGGGGGGGGGGGGGGGGGGGG 51
Db 10 GGGGGGGGGGGGGGGGGGGGGLGVLGGLISGAGGGGGGGGGGGGGGGGGGG 55

RESULT 14
US-09-157-349-9
; Sequence 9, Application US/09157349
; Patent No. 6068990
; GENERAL INFORMATION:
; APPLICANT: SHINTANI, Yasushi
; APPLICANT: NISHI, Kazuhiro
; APPLICANT: KAWAMOTO, Tomohiro
; TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION
; TITLE OF INVENTION: AND USE
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP
; STREET: 130 Water Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/157,349
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/835,099
; FILING DATE:
```

```
/ APPLICATION NUMBER: 97105508.2
/ FILING DATE: 03-APR-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Resnick, David S
/ REGISTRATION NUMBER: 34,235
/ REFERENCE/DOCKET NUMBER: 47342
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617-523-3400
/ TELEFAX: 617-523-6440
/ TELEX:
/ INFORMATION FOR SEQ ID NO: 9:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 268 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-157-349-9

Query Match 61.4%; Score 183; DB 3; Length 268;
Best Local Similarity 73.9%; Pred. No. 6.9e-11;
Matches 34; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Qy 6 GGGGGGGGGGGGGGGGGYGRKRRRRGGGGGGGGGGGGGGGGGG 51
Db 10 GGGGGGGGGGGGGGGGLGNVLGGLISGAGGGGGGGGGGGGGGG 55

RESULT 15
US-09-832-297A-12
/ Sequence 12, Application US/09832297A
/ Patent No. 6652836
/ GENERAL INFORMATION:
/ APPLICANT: FluoroProbe, Inc.
/ APPLICANT: LUIKEN, George A.
/ TITLE OF INVENTION: METHOD FOR VIEWING TUMOR TISSUE LOCATED WITHIN A BODY CAVITY
/ FILE REFERENCE: FLUOR1120-2
/ CURRENT APPLICATION NUMBER: US/09/832,297A
/ CURRENT FILING DATE: 2001-04-09
/ PRIOR APPLICATION NUMBER: US 09/362,805
/ PRIOR FILING DATE: 1999-07-28
/ PRIOR APPLICATION NUMBER: US 09/173,190
/ PRIOR FILING DATE: 1998-10-15
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 12
/ LENGTH: 60
/ TYPE: PRT
/ ORGANISM: Artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Peptide linker moiety
/ FEATURE:
/ NAME/KEY: REPEAT
/ LOCATION: (3)..(6)
/ OTHER INFORMATION: Amino Acid at residue 3 could be repeated up to 3 times
/ FEATURE:
/ NAME/KEY: REPEAT
/ LOCATION: (3)..(57)
/ OTHER INFORMATION: Amino Acids at residues 3 to 7 could be repeated up to 10 times
/ US-09-832-297A-12

Query Match 60.9%; Score 181.5; DB 4; Length 60;
Best Local Similarity 66.0%; Pred. No. 2.8e-11;
Matches 33; Conservative 0; Mismatches 8; Indels 9; Gaps 1;

Qy 2 GGGGGGGGGGGGGGGGGYGRKRRRRGGGGGGGGGGGGGGGGGG 51
Db 3 GGGGGGGGGGGGGGGGGG-----GGGGGGGGGGGGGGGG 43

Search completed: September 7, 2005, 13:01:57
Job time : 29.5 secs
```



```

RESULT 10
US-09-850-948-29
; Sequence 29, Application US/09850948
; Publication No. US20030059770A1
; GENERAL INFORMATION:
; APPLICANT: Yang, Jianxin
; APPLICANT: An, Songzhu
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Associated With Cancer
; FILE REFERENCE: 018781-008300US
; CURRENT APPLICATION NUMBER: US/09/850,948
; CURRENT FILING DATE: 2001-05-07
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly-Gly
; OTHER INFORMATION: flexible linker
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-09-850-948-29

Query Match      76.7%; Score 228.5; DB 10; Length 200;
Best Local Similarity   80.4%; Pred.No. 1.1e-11;
Matches    41; Conservative     0; Mismatches    1; Indels    9; Gaps    1;

QY       1  GGCGGGGCGGGGGGGGGGYGRKKRQRRRRGCGGGGGGGGGGGGGGGGGGG 51
          |||||-----GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG
DB        1  GGCGGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 42


RESULT 11
US-10-160-354-4
; Sequence 4, Application US/10160354
; Publication No. US20030013107A1
; GENERAL INFORMATION:
; APPLICANT: Chu, Peter
; APPLICANT: Li, Congfen
; APPLICANT: Liao, X. Charlene
; APPLICANT: Pardo, Jorge
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Alpha 2 Intergrin: Modulators of Lymphocyte Activation
; FILE REFERENCE: 021044-001110US
; CURRENT APPLICATION NUMBER: US/10/160,354
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 60/296,819
; PRIOR FILING DATE: 2001-06-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-10-160-354-4

Query Match      76.7%; Score 228.5; DB 14; Length 200;
Best Local Similarity   80.4%; Pred.No. 1.1e-11;
Matches    41; Conservative     0; Mismatches    1; Indels    9; Gaps    1;

QY       1  GGCGGGGCGGGGGGGGGGYGRKKRQRRRRGCGGGGGGGGGGGGGGGGG 51

```

[illegible]

RESULT 12

```

US-10-026-331B-21
; Sequence 21, Application US/10026331B
; Publication No. US20030027167A1
; GENERAL INFORMATION:
; APPLICANT: Hitoshi, Yasumichi
; APPLICANT: Demo, Susan
; APPLICANT: Jenkins, Yonchu
; APPLICANT: Regal Pharmaceuticals, Inc.
; TITLE OF INVENTION: MRE11: Modulation of Cellular Proliferation
; FILE REFERENCE: 021044-001310US
; CURRENT APPLICATION NUMBER: US/10/026,331B
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: US 60/309,737
; PRIOR FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly Gly
; OTHER INFORMATION: flexible linker
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly residues from position 6 to 200 may be present
; OTHER INFORMATION: or absent
US-10-026-331B-21

```

```
Query Match      76.7%; Score 228.5; DB 14; Length 200;
Best Local Similarity 80.4%; Pred. NO. 1.le-11;
Matches 41; Conservative 0; Mismatches 1; Indels 9;
```

RESULT 13

```

US-10-026-021-8
/ Sequence 8, Application US/10026021
/ Publication NO. US20030027756A1
/ GENERAL INFORMATION:
/ APPLICANT: Hitoshi, Yasumichi
/ APPLICANT: Demo, Susan
/ APPLICANT: Jenkins, Yonchu
/ APPLICANT: Rigal Pharmaceuticals, Inc.
/ TITLE OF INVENTION: SAK: Modulation of Cellular Proliferation for
/ TITLE OF INVENTION: Treatment of Cancer
/ FILE REFERENCE: 021044-001210US
/ CURRENT APPLICATION NUMBER: US/10/026,021
/ CURRENT FILING DATE: 2002-06-25
/ PRIOR APPLICATION NUMBER: US 60/309,632
/ PRIOR FILING DATE: 2001-08-01
/ NUMBER OF SEQ ID NOS: 8
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 8
/ LENGTH: 200
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:flexible linker
/ FEATURE:
/ NAME/KEY: MOD RES
/ LOCATION: (6)..(200)
/ OTHER INFORMATION: Gly at positions 6-200 may be present or absent
US-10-026-021-8

```

Query Match 76.7%; Score 228.5; DB 14; Length 200;
Best Local Similarity 80.4%; Pred. No. 1.1e-11;
Matches 41; Conservative 0; Mismatches 1; Indels 9;

[illegible]

RESULT 14

```

US-10-161-165-3
; Sequence 3, Application US/10161165
; Publication No. US20030027763A1
; GENERAL INFORMATION:
; APPLICANT: Bennett, Mark
; APPLICANT: Holland, Sacha
; APPLICANT: Rossi, Alex
; APPLICANT: Regal Pharmaceuticals, Incorporated
; TITLE OF INVENTION: CD43: Modulators of Mast Cell Degranulation
; FILE REFERENCE: 021044-001010US
; CURRENT APPLICATION NUMBER: US/10/161,165
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 60/296,801
; PRIOR FILING DATE: 2001-06-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly Gly t
; OTHER INFORMATION: flexible linker
; FEATURE:
; NAME/KEY: MOD_RBS
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly residues from position 6 to 200 may be pr
; OTHER INFORMATION: or absent
US-10-161-165-3

```

Query Match 76.7%; Score 228.5; DB 14; Length 200;
Best Local Similarity 80.4%; Pred. NO. 1.1e-11;
Matches 41; Conservative 0; Mismatches 1; Indels 9;

Qy 1 GGGGGGGGGGGGGGGGGGGGGGGRKRRRQRRGGGGGGGGGGGGGGGGGGG 51
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
pB 1 GG 42
 | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 15

```

US-10-160-663-3
; Sequence 3, Application US/10160663
; Publication NO. US20030040001A1
; GENERAL INFORMATION:
; APPLICANT: Demo, Susan
; APPLICANT: Hitoshi, Yasumichi
; APPLICANT: Pearaall, Denise
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: LETM1: Modulators of Cellular Proliferation
; FILE REFERENCES: 021044-000920US
; CURRENT APPLICATION NUMBER: US/10/160,663
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: US 60/296,817
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/347,970
; PRIOR FILING DATE: 2002-01-15
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 3
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2005, 12:47:59 ; Search time 24 Seconds
(without alignments)
204.461 Million cell updates/sec

Title: US-09-910-432-20
Perfect score: 298
Sequence: 1 GGGGGGGGGGGGGGGGGG.....GGGGGGGGGGGGGGGGGGGG 51

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:*
1: Pirl.*
2: Pirl.*
3: Pirl.*
4: Pirl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	221	74.2	221	2 T04592	glycine-rich cell
2	205	68.8	165	1 KNRZG1	glycine-rich cell
3	204.5	68.6	404	2 S54729	RNA-binding protei
4	203.5	68.3	165	2 T03583	glycine-rich RNA-b
5	203	68.1	136	2 T29282	hypothetical prote
6	202	67.8	396	2 T49109	glycine-rich prote
7	198	66.4	207	2 T07381	glycine-rich prote
8	197	66.1	440	2 S71795	transcription fact
9	196.5	65.9	285	2 T31503	hypothetical prote
10	196	65.8	255	2 B84777	hypothetical prote
11	194.5	65.3	168	1 S12312	glycine-rich RNA-b
12	194.5	65.3	291	1 S31415	glycine-rich prote
13	193.5	64.9	385	2 T20410	hypothetical prote
14	193	64.8	142	2 S12311	glycine-rich RNA-b
15	192	64.4	1226	2 T24045	hypothetical prote
16	191.5	64.3	106	2 P84797	hypothetical prote
17	191	64.1	167	2 S71779	glycine-rich RNA-b
18	190.5	63.9	384	1 A26099	glycine-rich cell
19	189.5	63.6	162	2 T10479	glycine-rich RNA-b
20	189.5	63.6	462	4 S33798	FUS/CHOP mutant fu
21	189.5	63.6	526	1 S33799	RNA-binding protei
22	189	63.4	252	1 S01821	glycine-rich prote
23	189	63.4	528	2 G02127	fus-like protein -
24	188	63.1	220	2 A44805	eggshell protein p
25	187	62.8	148	2 S41772	glycine-rich RNA-b
26	187	62.8	155	2 S20846	glycine-rich prote
27	187	62.8	214	1 KNNT25	glycine-rich prote
28	187	62.8	239	2 S49193	GCR 101 protein -
29	186.5	62.6	157	1 S14857	glycine-rich prote

30	186.5	62.6	199	2 S16063	acp-22 protein - Y
31	186.5	62.6	199	2 S32224	acp-22 protein - Y
32	186.5	62.6	271	2 S34666	glycine-rich prote
33	185.5	62.2	157	2 S04536	embryonic abundant
34	185.5	62.2	393	2 T20268	hypothetical prote
35	184.5	61.9	169	1 S38331	glycine-rich RNA-b
36	183	61.4	169	2 T10465	glycine-rich prote
37	183	61.4	268	1 CIHUL	calpain (EC 3.4.22
38	182.5	61.2	302	2 C84470	hypothetical prote
39	182	61.1	201	2 T00799	hypothetical prote
40	181	60.7	156	2 S41771	glycine-rich RNA-b
41	181	60.7	173	2 S53050	RNA binding protei
42	180.5	60.6	82	2 S19774	glycine-rich prote
43	180.5	60.6	408	2 S57483	glycine-rich protei
44	180.5	60.6	681	2 AB2155	hypothetical prote
45	180	60.4	161	2 S71453	glycine-rich RNA-b

ALIGNMENTS

RESULT 1

T04592
glycine-rich cell wall structural protein homolog F23E13.120 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004
C:Accession: T04592
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Jesse, T.
submitted to the Protein Sequence Database, March 1998
A:Reference number: 215378
A:Accession: T04592
A:Molecule type: DNA
A:Residues: 1-221 <BEV>
A:Cross-references: UNIPROT:O65514; EMBL:AL022141
A:Experimental source: cultivar Columbia; BAC clone F23E13
C:Genetics:
A:Map position: 4
A:Note: F23E13.120

Query Match 74.2%; Score 221; DB 2; Length 221;
Best Local Similarity 55.8%; Pred. No. 2.1e-11;
Matches 43; Conservative 0; Mismatches 8; Indels 26; Gaps 1;

Oy	1	GGGGGGG-----GGGGGGGGGGGGGVRKRRRRGGG	34
Db	98	GGGGGGGGGGGGGNGNGNHRNKSNGGGGGGGGGGGGSGRGRGGGG	157
Oy	35	GGGGGGGGGGGGGGGGG	51
Db	158	GGGGGGGGGGGGGGGGG	174

RESULT 2

KNRZG1
glycine-rich cell wall structural protein 1 precursor (clone lambda-313) - rice
C:Species: Oryza sativa (rice)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 09-Jul-2004
C:Accession: S13385
R:Lei, M.; Wu, R.
Plant Mol. Biol. 16, 187-198, 1991
A:Title: A novel glycine-rich cell wall protein gene in rice.
A:Reference number: S13385; MUID:91370862; PMID:1716496
A:Accession: S13385
A:Molecule type: DNA
A:Residues: 1-165 <LEI>
A:Cross-references: UNIPROT:P25074; EMBL:X53596; NID:g20246; PIDN:CAA37665.1; PID:g20247
C:Genetics:
A:Gene: grp-1
C:Superfamily: glycine-rich cell wall structural protein 1
C:Keywords: cell wall; duplication; structural protein
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-165/Product: glycine-rich cell wall structural protein 1 #status predicted <MAT>
F:30-55/Region: repeat R1


```

Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 1 GGGGGGGGGGGGGGGGGGGGGRKKRRQRRRRGGGGGGGGGGGGGGGGGGGG 51
    |||||
Db 286 GGGGGGGGGGGGGGGGGGGGSGSGGYGSGMGKSGSGSGGGGGGGGGGGGG 336
    |||||

RESULT 7
T07381
glycine-rich protein Tfm5 - tomato
C:Species: Lycopersicon esculentum (tomato)
C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 09-Jul-2004
C:Accession: T07381
R:Santino, C.G.; Stanford, G.L.; Conner, T.W.
Plant Mol. Biol. 33, 405-416, 1997
A:Title: Developmental and transgenic analysis of two tomato fruit enhanced genes.
A:Reference number: Z16000; MUID:97201476; PMID:9049262
A:Accession: T07381
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-207 <SAN>
A:Cross-references: UNIPROT:Q43522; EMBL:X95262; NID:g1166449; PIDN:CAA64559.1; PID:g116
A:Experimental source: cultivar UC82b; fruit
C:Genetics:
A:Gene: Tfm5
C:Superfamily: hydroxyproline-rich glycoprotein

Query Match 66.4%; Score 198; DB 2; Length 207;
Best Local Similarity 70.6%; Pred. No. 1.4e-09;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
Qy 1 GGGGGGGGGGGGGGGGGGGGGRKKRRQRRRRGGGGGGGGGGGGGGGGGGGG 51
    |||||
Db 73 GSGGGGSGSGSGSGSGSGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGG 123
    |||||

RESULT 8
S71795
transcription factor CBF-2 - chicken
C:Species: Gallus gallus (chicken)
C>Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 16-Aug-2004
C:Accession: S71795
R:Yuasa, J.; Hirano, S.; Yanagata, M.; Noda, M.
Nature 382, 632-635, 1996
A:Title: Visual projection map specified by topographic expression of transcription fac
A:Reference number: S71794; MUID:96338226; PMID:8757134
A:Accession: S71795
A>Status: preliminary; nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-440 <YUA>
A:Cross-references: UNIPROT:Q98937; EMBL:U47276; NID:g1546783; PIDN:AAB08467.1; PID:g154
A:Superfamily: fork head DNA-binding domain homology
F:143-234/Domain: fork head DNA-binding domain homology <FHD>

Query Match 66.1%; Score 197; DB 2; Length 440;
Best Local Similarity 63.9%; Pred. No. 2.8e-09;
Matches 39; Conservative 1; Mismatches 11; Indels 10; Gaps 2;
Qy 1 GGGGGGGGGGGGGGGGGGGGGRKKRR---QRR-----RGGGGGGGGGGGGGGGGGG 50
    |||||
Db 77 GGGGGGGGGGGAGAGGAGGDPGGRPPRPPGPPQRAAAGGGGAGAGGGGGGGGGGGGG 136
    |||||
Qy 51 G 51
    |
Db 137 G 137

RESULT 9
T31503
hypothetical protein Y116A8C.35 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T31503

```

```

R:McMurray, A.
submitted to the EMBL Data Library, October 1999
A:Reference number: Z21041
A:Accession: T31503
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-285 <WIL>
A:Cross-references: UNIPROT:Q9U2U0; EMBL:AL117204; PIDN:CAB55137.1; CESP:Y116A8C.35
A:Experimental source: clone Y116A8C
C:Genetics:
A:Gene: CESP:Y116A8C.35
A:Introns: 31/3; 143/2

```

```

Query Match 65.9%; Score 196.5; DB 2; Length 285;
Best Local Similarity 73.1%; Pred. No. 2.2e-09;
Matches 38; Conservative 1; Mismatches 12; Indels 1; Gaps 1;
Qy 1 GGGGGGGGGGGGGGGGGGGGGRKKRRQRRRRGGGGGGGGGGGGGGGGGGGG 51
    |||||
Db 213 GSGGGGGGGGGGGGGGGGGGGRDRGGGGGGGGGGGGGGGGGGGGGGGG 264
    |||||

```

```

RESULT 10
B84777
hypothetical protein At2g36120 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: B84777
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: B84777
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-255 <STO>
A:Cross-references: UNIPROT:Q9SIH2; GB:AE002093; NID:94678224; PIDN:AAD26969.1; GSPDB:GN
C:Genetics:
A:Map position: 2
C:Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology

```

```

Query Match 65.8%; Score 196; DB 2; Length 255;
Best Local Similarity 68.4%; Pred. No. 2.3e-09;
Matches 39; Conservative 1; Mismatches 11; Indels 6; Gaps 2;
Qy 1 GGGGGGGGGGGGGGGGGGG--GGGYRKRQRRRRRGGGGGG-----CGGGGGGGGGGGGG 51
    |||||
Db 125 GGGGGGAGGGGGGGGGGAHGGGYGGGGAGAGAGGGYGGGAGGGGGGGGGGGGGGGGGGG 181
    |||||

```

```

RESULT 11
S12312
glycine-rich RNA-binding protein (clone S2) - sorghum
C:Species: Sorghum bicolor (sorghum)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C:Accession: S12312
R:Cretin, C.; Puigdomenech, P.
Plant Mol. Biol. 15, 783-785, 1990
A:Title: Glycine-rich RNA-binding proteins from Sorghum vulgare.
A:Reference number: S12311; MUID:91346715; PMID:1715211
A:Accession: S12312
A:Molecule type: mRNA
A:Residues: 1-168 <CRE>
A:Cross-references: UNIPROT:Q99070; EMBL:X57662; NID:g21624; PIDN:CAA40862.1; PID:g21625
A>Note: in the authors' translation two additional Gly are shown after 110-Gly
C:Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology
C:Keywords: GTP binding
F:9-76/Domain: ribonucleoprotein repeat homology <RRM1>

```

```

Query Match 65.3%; Score 194.5; DB 1; Length 168;

```

```
Best Local Similarity 63.1%; Pred. No. 2.2e-09;
Matches 41; Conservative 0; Mismatches 5; Indels 19; Gaps 3;

Qy 1 GGGGGGGGGGGG-----GGGGGGRKRRRRGGGGGGGGGGGGGGG-----47
Db 89 GGGGGGGGGGGGGRGGGGGGGGGGG-----GRRGGGGGGGGGGGGGGR 143
Qy 48 -GGGG 51
Db 144 EGGGG 148

RESULT 12
S1415
glycine-rich protein GRP22 - rape
C/Species: Brassica napus (rape)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: S1415
R/Bergerson, D.; Boivin, R.; Baszczyński, C.L.; Bellemare, G.
submitted to the EMBL Data Library, August 1992
A/Description: Characterization and expression of a gene family encoding glycine-rich pr
A/Reference number: S1415
A/Accession: S1415
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-291 <BER>
A/Cross-references: UNIPROT:Q39337; EMBL:Z15045; NID:gl7820; PIDN:CAA78762.1; PID:gl7821
C/Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 65.3%; Score 194.5; DB 1; Length 291;
Best Local Similarity 65.0%; Pred. No. 3.3e-09;
Matches 39; Conservative 0; Mismatches 12; Indels 9; Gaps 2;

Qy 1 GGGGGGGGGGGGGG-----GGYGRKRRRRRRGGGG-----GGGGGGGGGGGGGGG 51
Db 207 GGGGGGGGGGGGGGAGAGHGGYGGGGGGGGGGGGGGGGGGGGGGGGGGGG 266

RESULT 13
T20410
hypothetical protein E02A10.2 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T20410
R/Thomas, K.
submitted to the EMBL Data Library, October 1996
A/Reference number: Z19271
A/Accession: T20410
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-385 <WIL>
A/Cross-references: UNIPROT:Q93424; EMBL:Z81053; PIDN:CAB02877.1; GSPDB:GN00023; CESP:E0
A/Experimental source: clone E02A10
C/Genetics:
A/Gene: CESP:E02A10.2
A/Map position: 5
A/Introns: 32/1; 72/1; 85/1; 122/1; 133/1; 220/3

Query Match 64.9%; Score 193.5; DB 2; Length 385;
Best Local Similarity 72.5%; Pred. No. 4.8e-09;
Matches 37; Conservative 0; Mismatches 5; Indels 9; Gaps 1;

Qy 1 GGGGGGGGGGGGGGGGGGGGGYGRKRRRRRRGGGGGGGGGGGGGGGGGGGGG 51
Db 136 GGGGGGGGGGGGGGGGGGGGGG-----GGGGGGGGGGGGGGGGGGGGG 177

RESULT 14
S12311
glycine-rich RNA-binding protein (clone S1) - sorghum (fragment)
C/Species: Sorghum bicolor (sorghum)
C/Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C/Accession: S12311
```

```
R/Cretin, C.; Puigdomenech, P.
Plant Mol. Biol. 15, 783-785, 1990
A/Title: Glycine-rich RNA-binding proteins from Sorghum vulgare.
A/Reference number: S12311; MUID:91346715; PMID:1715211
A/Accession: S12311
A/Molecule type: mRNA
A/Residues: 1-142 <CRE>
A/Cross-references: UNIPROT:Q99069; EMBL:X57663; NID:q21622; PIDN:CAA40863.1; PID:g21623
A/Note: the authors translated the codon ACG for residue 37 as Ser and CGC for residue 50
A/Note: the authors did not translate the codons for residues 92 and 93
C/Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology
C/Keywords: RNA binding
F;1-55/Domain: ribonucleoprotein repeat homology (fragment) <RRM1>

Query Match 64.8%; Score 193; DB 2; Length 142;
Best Local Similarity 70.9%; Pred. No. 2.6e-09;
Matches 39; Conservative 0; Mismatches 6; Indels 10; Gaps 2;

Qy 1 GGGGGGGGGGGGGGGGGGGGGYGRKRRRRRRGGGGGGGGGGGGGGG-----GGGG 51
Db 67 GGRGGGGGGGGYGGGGGGGGG-----RRDGGGGGGYGGGGGGGGGGGGG 115

RESULT 15
T24045
hypothetical protein R08B4.1 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T24045
R/White, S.
submitted to the EMBL Data Library, November 1995
A/Reference number: Z19834
A/Accession: T24045
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-1226 <WIL>
A/Cross-references: UNIPROT:Q21835; EMBL:Z68008; PIDN:CAA92000.1; GSPDB:GN00028; CESP:R0
A/Experimental source: clone R08B4
C/Genetics:
A/Gene: CESP:R08B4.1
A/Map position: X
A/Introns: 53/2; 113/1; 152/3; 204/3; 272/1; 354/1; 389/3; 714/3; 839/3; 877/1; 960/3; 1

Query Match 64.4%; Score 192; DB 2; Length 1226;
Best Local Similarity 66.7%; Pred. No. 1.4e-08;
Matches 34; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1 GGGGGGGGGGGGGGGGGGGGGYGRKRRRRRRGGGGGGGGGGGGGGGGGGGGG 51
Db 857 GCGNGGGGGGGGGGGGGGGGGGGGSGGSGGSGGSGGSGGSGGSGGSGGNGGG 907

Search completed: September 7, 2005, 13:00:51
Job time : 24 secs
```


Result No.	Query #			DB	ID	Description
	Score	Match	Length			
1	234	78.5	321	2	Q69XV3	Q69XV3 oryza sativ
2	221	74.2	221	2	Q65514	Q65514 arabidopsis
3	214.5	72.0	199	2	Q62PE6	Q62PE6 burkholderi
4	209.5	70.3	329	2	Q85215	Q85215 oryza sativ
5	205	68.8	165	1	GRP1 ORYSA	P25074 oryza sativ
6	204.5	68.6	175	2	Q9L5N6	Q9L5N6 arabidopsis
7	203.5	68.3	185	2	Q24184	Q24184 oryza sativ
8	203.5	68.3	551	2	Q75HG8	Q75HG8 oryza sativ
9	203	68.1	136	2	Q18444	Q18444 caenorhabdi
10	202	67.8	396	2	Q65450	Q65450 arabidopsis
11	200.5	67.3	535	2	Q942Q2	Q942Q2 oryza sativ
12	200.5	67.3	698	2	Q8QXK8	Q8QXK8 ectocarpus
13	199	66.8	165	2	Q95UX2	Q95UX2 drosophila
14	199	66.8	344	1	FBRL DROME	Q9W1V3 drosophila
15	199	66.8	345	1	FBZL DROER	Q811F4 drosophila
16	198.5	66.6	399	1	CAZL_DROME	Q27294 drosophila
17	198	66.4	207	2	Q43522	Q43522 lycopersico
18	197	66.1	155	2	Q9X8L4	Q9X8L4 picea glauc
19	197	66.1	239	2	Q69T79	Q69T79 oryza sativ
20	197	66.1	440	1	FXGA CHICK	Q98937 gallus gall
21	197	66.1	592	2	Q90W75	Q90W75 oncorhynch
22	196.5	65.9	161	2	Q22385	Q22385 oryza sativ
23	196.5	65.9	285	2	Q9U2U0	Q9U2U0 caenorhabdi
24	196.5	65.9	422	2	Q6Z142	Q6Z142 oryza sativ
25	196	65.8	161	2	Q95UW6	Q95UW6 drosophila
26	196	65.8	255	2	Q95IH2	Q95IH2 arabidopsis
27	196	65.8	301	2	Q15647	Q15647 plasmodium
28	195.5	65.6	192	2	Q92P87	Q92P87 rhizobium m
29	195	65.4	344	2	Q42403	Q42403 gallus gall
30	194.5	65.3	163	2	Q95UX4	Q95UX4 drosophila
31	194.5	65.3	165	2	Q95NR6	Q95NR6 drosophila

RESULT 4

[illegible]

RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celnikier S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shy S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.; of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;
RX MEDLINE=22426066; PubMed=12537559;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celnikier S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [4]
RP SEQUENCE OF 5-140 FROM N.A.
RX MEDLINE=87230988; PubMed=2884623;
RA Flavell A.J., Dyson J., Ish-Horowitz D.;
RT "A novel GC-rich dispersed repeat sequence in Drosophila
RT melanogaster.";
RL Nucleic Acids Res. 15:4035-4048(1987).
CC -!- FUNCTION: Fibrillarlin is a component of a nucleolar small nuclear
CC ribonucleoprotein particle thought to participate in the first
CC step in processing preribosomal RNA. It is associated with the U3,
CC U8 and U13 small nuclear RNAs (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear; fibrillar region of the nucleolus
CC (By similarity).
CC -!- PTM: By homology to other fibrillarins, some or all of the N-
CC terminal domain arginines are modified to asymmetric
CC dimethylarginine (DMA) (By similarity).
CC -!- SIMILARITY: Belongs to the fibrillarlin family.
CC -!- CAUTION: Ref.4 sequence differs from that shown due to erroneous
CC gene model prediction and to a frameshift in position 8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF003459; AAF46950.1; -;
DR EMBL; EF001738; AAN71493.1; -;
DR EMBL; X05285; CAA28903.1; ALT_FRAME.
DR HSSP; O57811; 1GBA.
DR IntAct; Q9W1V3; -;
DR FlyBase; FBgn003062; Fib.
DR InterPro; IPR000692; Fibrillarlin.
DR Pfam; PF01269; Fibrillarlin; 1.
DR PRINTS; PRO0052; FIBRILLARIN.
DR ProDom; PD004637; Fibrillarlin; 1.
DR PROSITE; PS00566; FIBRILLARIN; 1.
KW Methylation; Nuclear protein; Ribonucleoprotein; RNA-binding;
KW rRNA processing.
FT DOMAIN 9 108 Gly/DNA-rich.
FT MOD_RES 9 9 Asymmetric dimethylarginine (By
FT similarity).
FT MOD_RES 23 23 Asymmetric dimethylarginine (By
FT similarity).
FT MOD_RES 25 25 Asymmetric dimethylarginine (By
FT similarity).
FT MOD_RES 40 40 Asymmetric dimethylarginine (By
FT similarity).
FT MOD_RES 42 42 Asymmetric dimethylarginine (By
FT similarity).
FT MOD_RES 48 48 Asymmetric dimethylarginine (By
FT similarity).
FT MOD_RES 51 51 Asymmetric dimethylarginine (Probable).
FT MOD_RES 58 58 Asymmetric dimethylarginine (Probable).
FT

FT MOD_RES 63 63 Asymmetric dimethylarginine (Probable).
FT MOD_RES 71 71 Asymmetric dimethylarginine (Probable).
FT MOD_RES 77 77 Asymmetric dimethylarginine (Probable).
FT MOD_RES 83 83 Asymmetric dimethylarginine (Probable).
FT MOD_RES 88 88 Asymmetric dimethylarginine (Probable).
FT MOD_RES 93 93 Asymmetric dimethylarginine (Probable).
FT MOD_RES 98 98 Asymmetric dimethylarginine (Probable).
FT CONFLICT 55 67 Missing (in Ref. 3).
FT CONFLICT 134 134 R -> G (in Ref. 3).
SQ SEQUENCE 344 AA; 34637 MW; 58B536FACAE01D6 CRC64;
Query Match 66.8%; Score 199; DB 1; Length 344;
Best Local Similarity 65.1%; Pred. No. 4.2e-09;
Matches 41; Conservative 1; Mismatches 5; Indels 16; Gaps 3;
QY 1 GGGGGGGGGG---GGGGGGGGGGVGRKKRRRRRRGGG-----GGGGGGGGGGGG 48
DB 12 GGGGGGGGGGFRGCGGGGGGGGGGFG---GGRGRRGGGDRGGRGGGGRGGGG 67
QY 49 GGG 51
DB 68 GGG 70
RESULT 15
FBRL_DROER
ID FBRL_DROER STANDARD; PRT; 345 AA.
AC Q811F4;
DT 25-JAN-2005 (Rel. 46, Created)
DT 25-JAN-2005 (Rel. 46, Last sequence update)
DE 25-JAN-2005 (Rel. 46, Last annotation update)
DE Fibrillarlin.
GN Name=Fib;
OS Drosophila erecta (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriodes; Drosophilidae; Drosophila.
OX NCBI_TaxID=7220;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426072; PubMed=12537575;
RA Bergman C.M., Pfeiffer B.D., Rincon-Limas D.E., Hoskins R.A.,
RA Gnirke A., Mungall C.J., Wang A.M., Kronmiller B., Pacleb J.M.,
RA Park S., Stapleton M., Wan K.H., George R.A., de Jong P.J., Botas J.,
RA Rubin G.M., Celnikier S.E.;
RT "Assessing the impact of comparative genomic sequence data on the
RT functional annotation of the Drosophila genome.";
RL Genome Biol. 3:RESEARCH0086.1-RESEARCH0086.20(2002).
CC -!- FUNCTION: Fibrillarlin is a component of a nucleolar small nuclear
CC ribonucleoprotein particle thought to participate in the first
CC step in processing preribosomal RNA. It is associated with the U3,
CC U8 and U13 small nuclear RNAs (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear; fibrillar region of the nucleolus
CC (By similarity).
CC -!- PTM: By homology to other fibrillarins, some or all of the N-
CC terminal domain arginines are modified to asymmetric
CC dimethylarginine (DMA) (By similarity).
CC -!- SIMILARITY: Belongs to the fibrillarlin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AY190941; AAO1021.1; -;
DR HSSP; O57811; 1GBA.
DR FlyBase; FBgn0064623; DerefFib.
DR InterPro; IPR000692; Fibrillarlin.
DR Pfam; PF01269; Fibrillarlin; 1.
DR PRINTS; PRO0052; FIBRILLARIN.
DR

```
DR ProDom: PD004637; Fibrillarlin; 1.
DR PROSITE; PS00566; FIBRILLARIN; 1.
KW Methylation; Nuclear protein; Ribonucleoprotein; RNA-binding;
KW rRNA processing.
FT DOMAIN 9 109 Gly/DMA-rich.
FT MOD_RES 9 9 Asymmetric dimethylarginine (By
FT FT similarity).
FT MOD_RES 23 23 Asymmetric dimethylarginine (By
FT FT similarity).
FT MOD_RES 25 25 Asymmetric dimethylarginine (By
FT FT similarity).
FT MOD_RES 41 41 Asymmetric dimethylarginine (By
FT FT similarity).
FT MOD_RES 43 43 Asymmetric dimethylarginine (By
FT FT similarity).
FT MOD_RES 49 49 Asymmetric dimethylarginine (By
FT FT similarity).
FT MOD_RES 52 52 Asymmetric dimethylarginine (Probable).
FT MOD_RES 59 59 Asymmetric dimethylarginine (Probable).
FT MOD_RES 64 64 Asymmetric dimethylarginine (Probable).
FT MOD_RES 72 72 Asymmetric dimethylarginine (Probable).
FT MOD_RES 78 78 Asymmetric dimethylarginine (Probable).
FT MOD_RES 84 84 Asymmetric dimethylarginine (Probable).
FT MOD_RES 89 89 Asymmetric dimethylarginine (Probable).
FT MOD_RES 94 94 Asymmetric dimethylarginine (Probable).
FT MOD_RES 99 99 Asymmetric dimethylarginine (Probable).
SQ SEQUENCE 345 AA; 34709 MW; F05DAB01491AE66F CRC64;

Query Match 66.8%; Score 199; DB 1; Length 345;
Best Local Similarity 65.1%; Pred. No. 4.3e-09;
Matches 41; Conservative 1; Mismatches 5; Indels 16; Gaps 3;

QY 1 GGGGGGGGGG----GGGGGGGGGGYGRKKRRQRRRGGG-----GGGGGGGGGGGGG 48
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 13 GGGGGGGGGGFRGGGGGGGGGGGFGF----GGRGRGGGGDRGGRGFRGGGGGGGGG 68
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 49 GGG 51
    |||
Db 69 GGG 71
```

Search completed: September 7, 2005, 12:59:58
Job time : 113.5 secs

This Page Blank (usp10)

XX 02-JAN-2003.
 XX 18-JUN-2002; 2002EP-00013517.
 XX 18-JUN-2001; 2001JP-00246789.
 XX (NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.
 XX (ADSC-) CENT ADVANCED SCI & TECHNOLOGY INCUBATIO.
 XX Suwa M, Asai K, Akiyama Y, Aburatani H;
 XX WPI; 2003-315783/31.
 XX N-FSDB; ADC87060.
 XX New polynucleotide, useful for preparing a composition for treating a
 PT patient in need of increased or suppressed activity or expression of the
 PT guanosine triphosphate-binding protein coupled receptor.
 XX Claim 2; SEQ ID NO 1514; 28pp; English.
 XX The invention relates to a novel polynucleotide encoding a guanosine
 CC triphosphate-binding protein coupled receptor (GPCR). A polynucleotide of
 CC the invention may have a use in gene therapy. The polynucleotide and
 CC polypeptide are useful for preparing a composition for treating a patient
 CC in need of increased or suppressed activity or expression of the
 CC guanosine triphosphate-binding protein coupled receptor. The protein
 CC sequences shown in ADC85549-ADC87617 represent GPCR's of the invention.
 XX Sequence 920 AA;
 SQ

Query Match 77.3%; Score 230.5; DB 7; Length 920;
 Best Local Similarity 80.4%; Pred. No. 7.6e-14;
 Matches 41; Conservative 0; Mismatches 1; Indels 9; Gaps 1;
 Qy 1 GGGGGGGGGGGGGGGGGGGGGRRRRRRGGGGGGGGGGGGGGGGGGGGG 51
 Db 245 GGGGGGGGGGGGGGGGGGGGGXG-----GGGGGGGGGGGGGGGGGGG 286

RESULT 4
 ADG88556
 ID ADG88556 standard; protein; 200 AA.
 XX AC
 XX ADG88556;
 XX 11-MAR-2004 (first entry)
 XX Poly Gly flexible linker.
 XX Human; angiogenesis; contactin associated protein 3; CASPR3;
 KW gene therapy.
 KW Unidentified.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 6..200
 FT /note= "Gly residues from position 6 to 200 may be
 FT present or absent"
 XX
 XX US2003176333-A1.
 XX 18-SEP-2003.
 XX 18-MAR-2002; 2002US-00100818.
 XX 18-MAR-2002; 2002US-00100818.
 XX (RIGE-) RIGEL PHARM INC.
 XX Lorens JB, Xu W, Bogenberger J;
 XX WPI; 2003-843828/78.

XX Methods for identifying compounds which regulate angiogenesis via
 PT modulation of contactin associated protein 3 (CASPR3) for the used in
 PT diagnosis and therapy of diseases related to angiogenesis.
 XX
 XX Disclosure; SEQ ID NO 14; 61pp; English.
 XX The present invention relates to a method of identifying a compound that
 CC regulates angiogenesis via the modulation of contactin associated protein
 CC 3 (CASPR3). The method involves contacting the compound with a CASPR3
 CC polypeptide and determining the functional effect of the compound upon
 CC the CASPR3 polypeptide. The CASPR3 nucleic acid and encoded protein is
 CC useful as a drug target for anti-angiogenic therapies. The invention is
 CC also useful in gene therapy. The present sequence is poly Gly flexible
 CC linker. This sequence is used to illustrate the method of the invention.
 XX Sequence 200 AA;
 SQ

Query Match 77.0%; Score 229.5; DB 7; Length 200;
 Best Local Similarity 80.4%; Pred. No. 2.8e-14;
 Matches 41; Conservative 0; Mismatches 1; Indels 9; Gaps 1;
 Qy 1 GGGGGGGGGGGGGGGGGGGGGRRRRRRGGGGGGGGGGGGGGGGGGGGG 51
 Db 1 GGGGGGGGGGGGGGGGGGGGG-----GGGGGGGGGGGGGGGGGGG 42

RESULT 5
 ADJ84541
 ID ADJ84541 standard; protein; 200 AA.
 XX AC ADJ84541;
 XX 20-MAY-2004 (first entry)
 XX T2R G-protein coupled receptor related linker seq id 94.
 XX taste transduction; G-protein coupled receptor; T2R; taste signaling;
 KW CAMP level; cGMP level; IP3 level; Ca(2+) level; electrical activity;
 KW human; G-protein coupled receptor; receptor; linker.
 XX Synthetic.
 XX US2004038312-A1.
 XX 26-FEB-2004.
 XX 10-FEB-2003; 2003US-00364861.
 XX 10-SEP-1999; 99US-00393634.
 XX (ZUKE/) ZUKER C S.
 XX (ADLE/) ADLER J E.
 XX (HOON/) HOON M.
 XX (RYBA/) RYBA N.
 XX (MUEL/) MUELLER K.
 XX Zuker CS, Adler JE, Hoon M, Ryba N, Mueller K;
 XX WPI; 2004-203221/19.
 XX Novel isolated taste transduction G-protein coupled receptor e.g., T2R
 PT useful for identifying compound that modulates taste signaling in taste
 PT cells.
 XX Disclosure; SEQ ID NO 94; 121pp; English.
 XX The invention describes an isolated taste transduction G-protein coupled
 CC receptor (I) e.g., T2R which is expressed in a taste cell, comprises
 CC greater than 60% sequence identity to a fully defined sequence of 335
 CC (S1), 333 (S2), 299 (S3), 310 (S4), 224 (S5), 77 (S6), 209 (S7), 266
 CC (S8), 300 (S9), 300 (S10), 155 (S11), 173 (S12), 94 (S13), 115 (S14), 68
 CC (S15), 126 (S16) and 180 (S17) amino acids as given in the specification.

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-ABL3051), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (AB57737-
CC AB572072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 399 AA;

Query Match 69.3%; Score 206.5; DB 4; Length 399;
Best Local Similarity 68.9%; Pred. No. 6.7e-12;
Matches 42; Conservative 0; Mismatches 2; Indels 17; Gaps 3;

QY 1 GGGGGGGG-----GGGGGGGGGGGRRDRRRRGGGGGG-----GGGGGGGGGG 48
DB 215 GGGGGGGGGGFGGRRGGGGGGGGGGGGGRRFD-----RGGGGGGGRRDRGGGGGGGG 269
QY 49 G 49
DB 270 G 270

RESULT 10
ABU22369
ID ABU22369 standard; protein; 188 AA.
XX
AC ABU22369,
XX
XX 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #7896.
XX
DE Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
XX Burkholderia mallei.
XX
XX WO200277183-A2.
XX
XX 03-OCT-2002.
XX
XX 21-MAR-2002; 2002WO-US009107.
XX
XX 21-MAR-2001; 2001US-00815242.
XX
XX 06-SEP-2001; 2001US-00948993.
XX
XX 23-OCT-2001; 2001US-0342923P.
XX
XX 08-FEB-2002; 2002US-00072851.
XX
XX 06-MAR-2002; 2002US-0362659P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlson KL, Zyskind JW;
XX Wall D, Trawick JD, Carr GU, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
XX
XX N-PSDB; ACA26239.
XX
XX New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
XX isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 50293; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
XX the 6213 antisense sequences given in the specification where expression
XX of the nucleic acid inhibits proliferation of a cell. Also included are:
XX (1) a vector comprising a promoter operably linked to the nucleic acid
XX encoding a polypeptide whose expression is inhibited by the antisense
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX polypeptide or its fragment whose expression is inhibited by the
XX antisense nucleic acid; (4) an antibody capable of specifically binding
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular

CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 188 AA;

Query Match 68.1%; Score 203; DB 6; Length 188;
Best Local Similarity 68.9%; Pred. No. 7.9e-12;
Matches 42; Conservative 0; Mismatches 9; Indels 10; Gaps 3;

QY 1 GGGGGGGGGGGG--GGGGGGGGRRDRRRRGGGG--GGGG-----GGGGGGGGGG 50
DB 116 GGGGGGGGGGGGDDGGGGGGGGGGGGGGRRDRRRRGGGGGGGGGGGGGGGG 175
QY 51 G 51
DB 176 G 176

RESULT 11
ABO59421
ID ABO59421 standard; protein; 208 AA.
XX
XX ABO59421;
XX
XX 29-JUL-2004 (first entry)
XX
XX Human genome derived single exon protein #5655.
XX
XX Human; gene expression; single exon probe; microarray;
XX alternative splicing event; genomic alteration.
XX
XX Homo sapiens.
XX
XX US2003194704-A1.
XX
XX 16-OCT-2003.
XX
XX 03-APR-2002; 2002US-00029386.
XX
XX 03-APR-2002; 2002US-00029386.
XX
XX (PENN/) PENN S G.
XX (RANK/) RANK D R.
XX (HANZ/) HANZEL D K.
XX
XX Penn SG, Rank DR, Hanzel DK;
XX WPI; 2004-119264/12.
XX
XX New human genome-derived single exon nucleic acid probes useful for human
XX gene expression analysis, for identifying or characterizing alternative
XX splicing events, for assessing genomic alterations or as tools for
XX surveying tissues.


```
OS Drosophila melanogaster.
XX WO200171042-A2.
XX
XX
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US009231.
XX
XX 23-MAR-2000; 2000US-0191637P.
XX
XX 11-JUL-2000; 2000US-00614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX
XX N-PSDB; ABL08832.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signaling and cell-cell
XX interactions.
XX
XX Disclosure; SEQ ID NO 20979; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
XX sequences (ABU01840-ABU16175) and the encoded proteins (ABB57737-
XX ABB72072). The sequence data for this patent did not form part of the
XX printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 344 AA;
SQ
Query Match 66.1%; Score 197; DB 4; Length 344;
Best Local Similarity 70.9%; Pred. No. 4.6e-11;
Matches 39; Conservative 0; Mismatches 12; Indels 4; Gaps 1;
QY 1 GGGGGG----GGGGGGGGGGGGRRDRRRRRGGGGGGGGGGGGGGGGGGGG 51
Db 16 GGGGGGPRGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 70
RESULT 14
AAG51723
XX AAG51723 standard; protein; 258 AA.
XX
XX AAG51723;
XX
XX 18-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 65674.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-00301439.
XX
XX 25-FEB-1999; 99US-0121825P.
XX
XX 05-MAR-1999; 99US-0123180P.
XX
XX 09-MAR-1999; 99US-0123548P.
XX
XX 23-MAR-1999; 99US-0125788P.
XX
XX 23-MAR-1999; 99US-0126264P.
XX
29-MAR-1999; 99US-0126785P.
01-APR-1999; 99US-0127462P.
06-APR-1999; 99US-0128234P.
08-APR-1999; 99US-0128714P.
16-APR-1999; 99US-0129845P.
21-APR-1999; 99US-0130077P.
23-APR-1999; 99US-0130449P.
23-APR-1999; 99US-0130510P.
28-APR-1999; 99US-0130891P.
30-APR-1999; 99US-0131449P.
04-MAY-1999; 99US-0132048P.
04-MAY-1999; 99US-0132407P.
05-MAY-1999; 99US-0132484P.
06-MAY-1999; 99US-0132485P.
06-MAY-1999; 99US-0132486P.
11-MAY-1999; 99US-0132863P.
11-MAY-1999; 99US-0134256P.
14-MAY-1999; 99US-0134218P.
14-MAY-1999; 99US-0134219P.
14-MAY-1999; 99US-0134221P.
18-MAY-1999; 99US-0134370P.
18-MAY-1999; 99US-0134768P.
19-MAY-1999; 99US-0134941P.
20-MAY-1999; 99US-0135124P.
21-MAY-1999; 99US-0135353P.
24-MAY-1999; 99US-0135629P.
25-MAY-1999; 99US-0136021P.
27-MAY-1999; 99US-0136392P.
28-MAY-1999; 99US-0136782P.
01-JUN-1999; 99US-0137222P.
03-JUN-1999; 99US-0137528P.
04-JUN-1999; 99US-0137502P.
07-JUN-1999; 99US-0137724P.
08-JUN-1999; 99US-0138094P.
10-JUN-1999; 99US-0138540P.
10-JUN-1999; 99US-0138847P.
14-JUN-1999; 99US-0139119P.
16-JUN-1999; 99US-0139452P.
16-JUN-1999; 99US-0139453P.
17-JUN-1999; 99US-0139453P.
18-JUN-1999; 99US-0139492P.
18-JUN-1999; 99US-0139454P.
18-JUN-1999; 99US-0139455P.
18-JUN-1999; 99US-0139456P.
18-JUN-1999; 99US-0139457P.
18-JUN-1999; 99US-0139458P.
18-JUN-1999; 99US-0139459P.
18-JUN-1999; 99US-0139460P.
18-JUN-1999; 99US-0139461P.
18-JUN-1999; 99US-0139462P.
18-JUN-1999; 99US-0139463P.
18-JUN-1999; 99US-0139750P.
18-JUN-1999; 99US-0139763P.
21-JUN-1999; 99US-0139817P.
22-JUN-1999; 99US-0139899P.
23-JUN-1999; 99US-0140353P.
23-JUN-1999; 99US-0140354P.
24-JUN-1999; 99US-0140695P.
28-JUN-1999; 99US-0140823P.
29-JUN-1999; 99US-0140991P.
30-JUN-1999; 99US-0141287P.
01-JUL-1999; 99US-0141842P.
02-JUL-1999; 99US-0142055P.
06-JUL-1999; 99US-0142390P.
08-JUL-1999; 99US-0142803P.
09-JUL-1999; 99US-0142920P.
12-JUL-1999; 99US-0142977P.
13-JUL-1999; 99US-0143542P.
14-JUL-1999; 99US-0143624P.
15-JUL-1999; 99US-0144005P.
16-JUL-1999; 99US-0144085P.
16-JUL-1999; 99US-0144086P.
```


PR 27-MAR-2002; 2002US-0367500P.
PR 08-APR-2002; 2002US-0370227P.
PR 10-MAY-2002; 2002US-0378950P.
PR 24-JUL-2002; 2002US-0398008P.
PR 09-AUG-2002; 2002US-0402131P.
PR 13-AUG-2002; 2002US-0402708P.
PR 18-SEP-2002; 2002US-0411355P.
PR 02-OCT-2002; 2002US-0414984P.
PR 11-OCT-2002; 2002US-0417611P.
PR 23-OCT-2002; 2002US-0420246P.
PR 05-NOV-2002; 2002US-0423623P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Haseltine WA;
XX
XX WPI; 2003-598501/56.
XX
XX New albumin fusion protein, useful for preparing a composition for
PT treating diabetes mellitus.
XX
XX
XX Example 24; SEQ ID NO 858; 1086pp; English.
XX
XX The invention relates to fusion proteins comprising human serum albumin
CC (ADH21530) and a therapeutic polypeptide such as a therapeutic protein,
CC antibody or peptide or their variants or fragments. The therapeutic
CC protein may be fused to the N-terminus, the C-terminus or both termini of
CC albumin via a linker. The albumin component of the fusion proteins
CC prolongs the shelf-life and the in vitro and vivo biological activity of
CC the proteins compared with those of the corresponding therapeutic
CC proteins on their own. The invention also relates to nucleic acids
CC encoding albumin fusion proteins, vectors and host cells comprising an
CC albumin fusion protein nucleic acid, compositions and kits comprising an
CC albumin fusion protein, the method of extending the shelf-life of a
CC therapeutic protein by fusion with albumin, and the treatment of disease
CC using an albumin fusion protein. The albumin fusion proteins may be used
CC in the treatment of metabolic/endocrine disorders, diabetes and diabetes-
CC related conditions. Specifically the albumin fusion proteins may be used
CC to treat type 1 and type 2 diabetes, hyperglycaemia, neural disorders
CC (especially neuropathy), retinopathy, cardiovascular disorders
CC (especially heart disease, renal disorders and obesity. The proteins may
CC also be used in a method of maintaining a basal glucose level in a
CC patient and in a method for losing weight. The present sequence is
CC related to the invention.
XX
XX SQ Sequence 38 AA;
Query Match 64.6%; Score 192.5; DB 7; Length 38;
Best Local Similarity 74.5%; Pred. No. 2.1e-11;
Matches 35; Conservative 0; Mismatches 1; Indels 11; Gaps 1;
QY 1 GGGGGGGGGGGGGGGGGGGGGRRRRRRGGGGGGGGGGGGGGG 47
Db 1 GGG 36

Search completed: September 7, 2005, 12:56:07
Job time : 68.5 secs


```
; REFERENCE/DOCKET NUMBER: 018781-000800US
;
; TELEPHONE INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1..97
; OTHER INFORMATION: /product= "OTHER"
; OTHER INFORMATION: /note="gly at positions 1-97 may be
; OTHER INFORMATION: present or absent"
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 105..201
; OTHER INFORMATION: /product= "OTHER"
; OTHER INFORMATION: /note="gly at positions 105-201 may be
; OTHER INFORMATION: present or absent"
; US-09-053-003-40
Query Match 77.0%; Score 229.5; DB 3; Length 201;
Best Local Similarity 80.4%; Pred.No.3.ee-15;
Matches 41; Conservative 0; Mismatches 1; Indels 9;

QY      1 GGGGGGGGGGGGGGGGGRDRRRRQQRGGGGGGGGGGGGGGGGGGGGGGGG
Db       1 GGGGGGGGGGGGGGGGGGGGGGG-----GGGGGGGGGGGGGGGGGGGG
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

[illegible]

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 7, 2005, 12:54:00 ; Search time 57 Seconds
(without alignment)
352.368 Million cell updates/sec

Title: US-09-910-432-19

Perfect score: 298

Sequence: 1 GGGGGGGGGGGGGGGGGG.....GGGGGGGGGGGGGGGGGG 51

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1774312 seqs, 393823214 residues

Total number of hits satisfying chosen parameters: 1774312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*

2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*

6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*

7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*

9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*

10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*

11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*

12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*

13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*

14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*

15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*

16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*

17: /cgn2_6/ptodata/1/pubpaa/US10E_PUBCOMB.pep.*

18: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*

19: /cgn2_6/ptodata/1/pubpaa/US11A_PUBCOMB.pep.*

20: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*

21: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*

22: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	298	100.0	51	10	US-09-910-432-19
2	277	93.0	51	10	US-09-910-432-20
3	230.5	77.3	920	15	US-10-292-798-1514
4	229.5	77.0	200	9	US-09-798-584-18
5	229.5	77.0	200	9	US-09-967-624-19
6	229.5	77.0	200	9	US-09-986-667-18
7	229.5	77.0	200	10	US-09-921-159-34
8	229.5	77.0	200	10	US-09-990-940-21
9	229.5	77.0	200	10	US-09-989-981A-13
10	229.5	77.0	200	10	US-09-850-948-29
11	229.5	77.0	200	14	US-10-160-354-4

12	229.5	77.0	200	14	US-10-026-331B-21	Sequence 21, Appl
13	229.5	77.0	200	14	US-10-026-021-8	Sequence 8, Appl
14	229.5	77.0	200	14	US-10-161-165-3	Sequence 3, Appl
15	229.5	77.0	200	14	US-10-160-663-3	Sequence 3, Appl
16	229.5	77.0	200	14	US-10-071-838-15	Sequence 15, Appl
17	229.5	77.0	200	14	US-10-094-417-25	Sequence 25, Appl
18	229.5	77.0	200	14	US-10-188-405-13	Sequence 13, Appl
19	229.5	77.0	200	14	US-10-273-575-29	Sequence 29, Appl
20	229.5	77.0	200	14	US-10-233-098-5	Sequence 5, Appl
21	229.5	77.0	200	14	US-10-245-850-3	Sequence 5, Appl
22	229.5	77.0	200	14	US-10-293-582-27	Sequence 27, Appl
23	229.5	77.0	200	14	US-10-100-818-14	Sequence 14, Appl
24	229.5	77.0	200	14	US-10-237-467-18	Sequence 18, Appl
25	229.5	77.0	200	14	US-10-321-204-54	Sequence 54, Appl
26	229.5	77.0	200	14	US-10-179-766-12	Sequence 12, Appl
27	229.5	77.0	200	14	US-10-123-568-4	Sequence 4, Appl
28	229.5	77.0	200	14	US-10-123-731-10	Sequence 10, Appl
29	229.5	77.0	200	15	US-10-308-393-53	Sequence 53, Appl
30	229.5	77.0	200	15	US-10-352-724-5	Sequence 5, Appl
31	229.5	77.0	200	15	US-10-328-916-57	Sequence 57, Appl
32	229.5	77.0	200	15	US-10-452-015-3	Sequence 3, Appl
33	229.5	77.0	200	15	US-10-364-861-94	Sequence 94, Appl
34	229.5	77.0	200	15	US-10-231-956A-521	Sequence 521, Appl
35	229.5	77.0	200	15	US-10-255-775-3	Sequence 3, Appl
36	229.5	77.0	200	15	US-10-459-190-24	Sequence 24, Appl
37	229.5	77.0	200	15	US-10-339-744-4	Sequence 4, Appl
38	229.5	77.0	200	16	US-10-649-400-7	Sequence 7, Appl
39	229.5	77.0	200	16	US-10-620-052A-78	Sequence 78, Appl
40	229.5	77.0	200	16	US-10-794-897-9	Sequence 9, Appl
41	229.5	77.0	200	17	US-10-616-403-7	Sequence 7, Appl
42	229.5	77.0	200	17	US-10-679-102-32	Sequence 32, Appl
43	229.5	77.0	200	17	US-10-696-909A-72	Sequence 72, Appl
44	229.5	77.0	200	18	US-10-491-545A-96	Sequence 96, Appl
45	229.5	77.0	200	18	US-10-805-075-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-09-910-432-19

; Sequence 19, Application US/09910432

; Publication No. US20030229034A1

; GENERAL INFORMATION:

; APPLICANT: Waugh, Jacob

; APPLICANT: Dake, Michael

; APPLICANT: Essentia Biosystems, Inc.

; TITLE OF INVENTION: Multi-Component Biological Transport Systems

; FILE REFERENCE: 020154-000110US

; CURRENT APPLICATION NUMBER: US/09/910,432

; PRIOR FILING DATE: 2001-07-20

; PRIOR APPLICATION NUMBER: US 60/220,244

; PRIOR FILING DATE: 2000-07-21

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 19

; TYPE: PRT

; LENGTH: 51

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:positively charged HIV-TAR fragment branching group

; OTHER INFORMATION: (efficiency group) attached to solid backbone

; FEATURE:

; NAME/KEY: MOD RES

; LOCATION: (1)..(20)

; OTHER INFORMATION: Gly at positions 1-20 may be present or absent

; FEATURE:

; NAME/KEY: MOD RES

; LOCATION: (32)..(51)

; OTHER INFORMATION: Gly at positions 32-51 may be present or absent

US-09-910-432-19


```

RESULT 6
US-09-398-667-18
; Sequence 18, Application US/09998667
; Patent No. US20020146747A1
; GENERAL INFORMATION:
; APPLICANT: Masuda, Esteban
; APPLICANT: Liao, X. Charlene
; APPLICANT: Zhao, Haoran
; APPLICANT: Chu, Peter
; APPLICANT: Pardo, Jorge
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: TRACs: Modulators of Lymphocyte Activation
; FILE REFERENCE: 021044-000600US
; CURRENT APPLICATION NUMBER: US/09/998,667
; CURRENT FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: US 60/282,432
; PRIOR FILING DATE: 2001-04-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 18
; LENGTH: 200
; TYPE: PRT

```

```

RESULT 8
US-09-990-940-21
; Sequence 21, Application US/09990940
; Publication No. US20030027252A1
; GENERAL INFORMATION:
; APPLICANT: Tian, Hui
; APPLICANT: Zhao, Jiagang
; APPLICANT: Chen, Jin-Long
; APPLICANT: Cutler, Gene
; APPLICANT: An, Songzhu
; APPLICANT: Dai, Kang
; APPLICANT: Gupte, Jámila S.
; APPLICANT: Tularik inc.
; TITLE OF INVENTION: No. US20030027252A1el Receptors
; FILE REFERENCE: 018781-007410US
; CURRENT APPLICATION NUMBER: US/09/990,940
; CURRENT FILING DATE: 2001-11-21

```

```

; PRIOR APPLICATION NUMBER: US 60/252,841
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: US 60/257,636
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 60/261,377
; PRIOR FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/279,554
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: US 60/280,696
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; NAME/KEY: MOD RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
; US-09-990-940-21

Query Match          77.0%; Score 229.5; DB 10; Length 200;
Best Local Similarity 80.4%; Pred. No. 6.7e-12;
Matches 41; Conservative 0; Mismatches 1; Indels 9; Gaps 1;

Qy 1 GGGGGGGGGGGGGGGGGGGGGRRRRRRGGGGGGGGGGGGGGGGGGGGGG 51
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 42
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
US-09-989-981A-13
; Sequence 13, Application US/09989981A
; Publication No. US20030049730A1
; GENERAL INFORMATION:
; APPLICANT: Hobbs, Helen H.
; APPLICANT: Shan, Bei
; APPLICANT: Barnes, Robert
; APPLICANT: Tian, Hui
; APPLICANT: Tularik Inc.
; APPLICANT: Board of Regents, The University of Texas System
; TITLE OF INVENTION: ABCG5 and ABCG8: Compositions and Methods of Use
; FILE REFERENCE: 018781-007320US
; CURRENT APPLICATION NUMBER: US/09/989,981A
; PRIOR FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 60/252,235
; PRIOR FILING DATE: 2000-11-20
; PRIOR APPLICATION NUMBER: US 60/253,645
; PRIOR FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; NAME/KEY: MOD RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
; US-09-989-981A-13

Query Match          77.0%; Score 229.5; DB 10; Length 200;
Best Local Similarity 80.4%; Pred. No. 6.7e-12;
Matches 41; Conservative 0; Mismatches 1; Indels 9; Gaps 1;

Qy 1 GGGGGGGGGGGGGGGGGGGGGRRRRRRGGGGGGGGGGGGGGGGGGGG 51
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 42
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 10
US-09-850-948-29
; Sequence 29, Application US/09850948
; Publication No. US20030059770A1
; GENERAL INFORMATION:
; APPLICANT: Yang, Jianxin
; APPLICANT: An, Songzhu
; APPLICANT: Tularik Inc.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Associated With Cancer
; FILE REFERENCE: 018781-008300US
; CURRENT APPLICATION NUMBER: US/09/850,948
; CURRENT FILING DATE: 2001-05-07
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:poly-Gly
; OTHER INFORMATION: flexible linker
; NAME/KEY: MOD RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
; US-09-850-948-29

Query Match          77.0%; Score 229.5; DB 10; Length 200;
Best Local Similarity 80.4%; Pred. No. 6.7e-12;
Matches 41; Conservative 0; Mismatches 1; Indels 9; Gaps 1;

Qy 1 GGGGGGGGGGGGGGGGGGGGGRRRRRRGGGGGGGGGGGGGGGGGGGG 51
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 42
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
US-10-160-354-4
; Sequence 4, Application US/10160354
; Publication No. US20030013107A1
; GENERAL INFORMATION:
; APPLICANT: Chu, Peter
; APPLICANT: Li, Congfen
; APPLICANT: Liao, X. Charlene
; APPLICANT: Pardo, Jorge
; APPLICANT: Rigel Pharmaceuticals, Incorporated
; TITLE OF INVENTION: Alpha 2 Intergrin: Modulators of Lymphocyte Activation
; FILE REFERENCE: 021044-001110US
; CURRENT APPLICATION NUMBER: US/10/160,354
; CURRENT FILING DATE: 2002-05-30
; PRIOR APPLICATION NUMBER: US 60/296,819
; PRIOR FILING DATE: 2001-06-07
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:flexible linker
; NAME/KEY: MOD RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly at positions 6-200 may be present or absent
; US-10-160-354-4

Query Match          77.0%; Score 229.5; DB 14; Length 200;
Best Local Similarity 80.4%; Pred. No. 6.7e-12;
Matches 41; Conservative 0; Mismatches 1; Indels 9; Gaps 1;

Qy 1 GGGGGGGGGGGGGGGGGGGGGRRRRRRGGGGGGGGGGGGGGGGGGGG 51
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 42
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
```



```

; OTHER INFORMATION: Description of Artificial Sequence:poly Gly tag
; OTHER INFORMATION: flexible linker
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (6)..(200)
; OTHER INFORMATION: Gly residues from position 6 to 200 may be present
; OTHER INFORMATION: or absent
US-10-160-663-3

Query Match      77.0%; Score 229.5; DB 14; Length 200;
Best Local Similarity 80.4%; Pred. No. 6.7e-12;
Matches 41; Conservative 0; Mismatches 1; Indels 9; Gaps 1;

QY      1 GGGGGGGGGGGGGGGGGGGGGRRRRGGGGGGGGGGGGGGGGGGGG 51
      ||||||||||||||||| | |||||||||||||||||
Db      1 GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 42
      ||||||||||||||||| | |||||||||||||||||

Search completed: September 7, 2005, 13:03:57
Job time : 58 secs

```

This Page Blank (uspio)

A:Accession: S54729
A:Molecule type: DNA
A:Residues: 1-404 <STO>
A:Cross-references: UNIPROT:Q27294; EMBL:L37083; NID:g567105; PIDN:AAC41563.1; PID:g567105
R:Stolow, D.T.; Haynes, S.R.
Nucleic Acids Res. 23, 835-843, 1995
A:Title: Cabeza, a *Drosophila* gene encoding a novel RNA binding protein, shares homology
A:Reference number: S54728; MUID:95223793; PMID:7708500
A:Accession: S54728
A>Status: nucleic acid sequence not shown

A;Residues: 1-106 <STO>
A;Cross-references: UNIPROT:O7XUP7; GB:AE002093; NID:g4895201; PIDN:AAD32788.1; GSPDB:G22
C;Genetics:
A;Gene: At2g37830
A;Map position: 2
C;Superfamily: Arabidopsis glycine-rich protein 3

Query Match 65.3%; Score 194.5; DB 2; Length 106;
Best Local Similarity 55.7%; Pred.No. 1.5e-09;
Matches 39; Conservative 0; Mismatches 12; Indels 19; Gaps 1;

QY 1 GGGGGGGGGGGGGGGGGGRD-----DRRRRRGGGGGGGGG 41
| | | | | | | | | | | | | | | | | | | | |
Db 31 GGGGGGGGGGGGGGGGGEDGGDEIGDGANGGFGGDGGGGGGGGG 90
| | | | | | | | | | | | | | | | | | | | |
QY 42 GGGGGGGGGG 51
| | | | | | | | | | | | | | | | | | | | |
Db 91 GGGGGGGGGG 100
| | | | | | | | | | | | | | | | | | | | |

RESULT 10
T03583
glycine-rich RNA-binding protein - rice
C;Species: Oryza sativa (rice)
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 09-Jul-2004
C;Accession: T03583
R;Lee, M.C.; Kim, C.S.; Yi, B.Y.; Lee, J.S.; Eun, M.Y.
submitted to the EMBL Data Library, June 1997
A;Description: Isolation and characterization of RNA-binding glycine rich protein of ric.
A;Reference number: Z14958
A;Accession: T03583
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-165 <LEE>
A;Cross-references: UNIPROT:O24184; EMBL:AF009411; NID:g2267592; PIDN:AAB63589.1; PID:g22
A;Experimental source: cv. Wilyang 23
C;Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology
F;9-76/Domain: ribonucleoprotein repeat homology <RRM>

Query Match 65.3%; Score 194.5; DB 2; Length 165;
Best Local Similarity 67.2%; Pred.No. 2.1e-09;
Matches 41; Conservative 0; Mismatches 5; Indels 15; Gaps 3;

QY 1 GGGGGGG---GGGGG-----GGGGGGGGRDRRRRRRRGGGGGGGGGGGGGGG 50
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 90 GGGGGGYQRGGGGYGGRGYGGGGGGYG----QRREGGYGGGGYGGGGGGYG 144
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 51 G 51
|
Db 145 G 145

RESULT 11
T49109
glycine-rich protein - Arabidopsis thaliana
N;Alternate names: protein AT4g22020
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C;Accession: T49109
B;Evans, M.; Wedler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mf.
submitted to the Protein Sequence Database, May 2000
A;Reference number: Z25016
A;Accession: T49109
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-396 <BEV>
A;Cross-references: UNIPROT:O65450; EMBL:AL022140; GSPDB:GN00062; ATSP:AT4g22020
A;Experimental source: cultivar Columbia; BAC clone F1N20
C;Genetics:
A;Gene: ATSP:AT4g22020
A;Map position: 4
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

